VU Research Portal

Typicality, physiological activity and concept identification.

Das-Smaal, E.A.; de Swart, J.H.

published in

Acta Psychologica
1981

DOI (link to publisher)

[10.1016/0001-6918\(81\)90036-6](https://doi.org/10.1016/0001-6918(81)90036-6)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Das-Smaal, E. A., & de Swart, J. H. (1981). Typicality, physiological activity and concept identification. *Acta Psychologica*, 47, 1-24. [https://doi.org/10.1016/0001-6918\(81\)90036-6](https://doi.org/10.1016/0001-6918(81)90036-6)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

TYPICALITY, PHYSIOLOGICAL ACTIVITY AND CONCEPT IDENTIFICATION *

E.A. DAS-SMAAL and J.H. DE SWART

Laboratory for Experimental Psychology, Free University, Amsterdam, The Netherlands

Accepted November 1979

The extent to which instances are good or poor examples of their categories (typicality) was varied in a concept identification (CI) task. Typicality was first established for the kind of artificial material traditionally used in CI tasks (experiment 1). This material was employed in a CI task (experiment 2) with a variety of dependent variables, including the galvanic skin response (GSR). To test the generality of the results, more realistic stimuli were employed (experiments 3 and 4). The results showed that typicality influenced performance on the CI task, that the GSR is primarily related to uncertainty reduction, that the findings with arbitrary materials are replicated with more meaningful materials, and that the multi-hypothesis sampling theory of Levine is supported by the findings.

Introduction

Most of the 'classical' studies of concept identification (CI) in the tradition of Bruner *et al.* (1956) can be criticized on several points. In a CI task the subject is presented with a series of stimuli, some of which are instances of the concept, and others of which are not. The subject is asked to categorize each stimulus. Subsequently feedback is given regarding the correct categorization. One point of criticism concerns the equivalence of instances as members of a category, and the discreteness of their attributes. In many of the traditional studies the discrete values of each dimension were chosen because of their discriminability, and the chosen values remained constant during the experiment (*e.g.* only one value of red, blue, *etc.* in the color dimension was used). However, in many everyday categories values are continuous rather than discrete (*e.g.* different variations of red, blue, *etc.* in the color dimension).

* Requests for reprints may be sent to E.A. Das-Smaal, Vakgroep Funktieeler and Methodenleer, Free University, De Boelelaan 1115, 1007 MC Amsterdam BTV, The Netherlands.

Moreover, not all instances are equally good examples of their categories. They may differ in exemplariness or typicality (e.g. Rosch 1973, 1975a, b; Smith *et al.* 1974). It is on the issue of typicality that the present study aims to adjust the traditional type of CI study to a new approach, started by Rosch. Rosch's (1973) theory of internal structure applies prototype theory of natural concepts. She suggested that best examples or prototypes form the focus of organization, which is surrounded by other instances of decreasing similarity to that core.

As a possible determining factor of internal category structure, Rosch (1973) suggested the principle of maximization of cue validity. That is, best examples of categories are those instances which have the most in common with other members of the same category, and share the least with contrasting categories. In this view, typicality is determined both by a within, and by a between category aspect. Regarding the between category aspect, evidence for influence of proximity on classification performance was found by Goldman and Homa (1977). Between category discriminability was manipulated by altering the amount of feature overlap, and low feature overlap resulted in better classification performance than high feature overlap. Cutting and Schatz (1976) found that in a card-sorting task typicality interacted with proximity of categories, with very long decision times for poor examples of semantically close categories. With respect to the study of CI performance also Archer's (1962) findings on obviousness are of interest. Obviousness was defined as the magnitude of the difference between two categories. Archer observed a better performance when relevant information was highly rather than slightly obvious; the reverse was found for irrelevant information. Typicality, however, also implies the within category aspect. Typicality has to do with variability of the values itself.

The purpose of the present study was fourfold. The first aim was to investigate the effect on CI performance of varying the typicality of dimensional values. Following Rosch's (1973) line of thought on the facilitating effect of prototypical information on learning, and from Archer's (1962) findings regarding obviousness, it was predicted that CI performance would be better when the relevant (or defining) values (R) were typical (typ) than when they are atypical (atyp). Furthermore, when the irrelevant (or nondefining) values (I) were atypical, CI performance would be better than when they were typical. A combination of both predictions resulted in the following expected order of increasing

difficulty in CI performance (using the above mentioned abbreviations; < and & stand for 'easier than' and 'combined with', respectively):

(R typ & I atyp) < (R typ & I typ) < (R atyp & I atyp) < (R atyp & I typ).

In a preliminary experiment ratings of typicality will be established for different dimensional values. These ratings will be used to test the above mentioned predictions.

The second aim of the present study was to elaborate further the relationship between the amount of information processing in cognitive tasks and autonomic physiological activity. In a CI task De Swart and Das-Smaal (1976, 1979a,b) found a positive relationship between GSR and the informational value of the feedback; GSR appeared to be lowest at blanks (which provided no information), higher at confirming feedback, and highest at disconfirming feedback. Furthermore, GSR at disconfirming feedback increased with an increasing number of confirmations directly preceding disconfirmations. Moreover, a relationship was found between subject certainty about the classification response and GSR, for each type of feedback. At confirming feedback smaller GSRs were found with high certainty than with low certainty, whereas at disconfirming feedback the reverse was found. At blanks, GSR was not influenced by certainty about classification. From these results De Swart and Das-Smaal (1979a) concluded that GSR primarily reflects uncertainty reduction, and not information processing activities during a CI task. The present study attempts to obtain further information on this issue. De Swart and Das-Smaal's interpretation prescribes (1) that GSR does not vary systematically among tasks in which different amounts of information processing activities are involved, but which include no uncertainty reduction about alternative hypotheses; (2) that GSR will vary in CI tasks as a function of uncertainty reduction but not as a function of the difficulty of the CI problem.

The third aim of the study was to investigate whether the results are influenced by stimulus material. Traditional material was compared with more 'real-life' material. In correspondence with classical CI studies, the first two experiments contained meaningless, artificial stimuli. These consisted of unconnected, arbitrarily arranged, separable dimensions. The next two experiments replicated the first ones with more meaningful, realistic material, which was made more unitary. This

was done because, on intuitive grounds, unitariness of dimensions seemed to correspond more to real-life.

According to Garner (1976), selective attention is easier with separable than with unitary dimensions. More specifically, in a CI task the irrelevant dimensions can be ignored more easily when the dimensions are separable, and this should facilitate learning. Posner (1964) also stated that separability of dimensions is of importance in classification tasks; several studies showed that irrelevant dimensions interfered with performance with unitary but not with separable stimuli. In the present study, therefore, it was predicted that performance would be influenced by stimulus material, and that CI would be faster with meaningless, separable stimuli than with meaningful, more unitary material.

Finally, the CI tasks of the present experiment offered the opportunity to replicate Levine's (1969) study, in which evidence for his multi-hypothesis sampling theory was found. According to this theory, the subject monitors a subset of hypotheses and uses one of these, the working hypothesis, as the basis for his response. By rejecting those hypotheses that are disconfirmed by the feedback, the subject reduces his subset until only the correct hypothesis remains. According to this view, the trial of the last error (TLE) is that trial at which the subject for the first time takes the correct hypothesis as his new working hypothesis. During the following trials, the subset is narrowed down to one, the correct hypothesis. Assuming that the latency of the response is a function of the number of hypotheses to be evaluated in the subset, there should be a decrease in latency from TLE to the trial at which a single hypothesis remains, the solution trial (ST), whereas beyond the ST the latencies should be constant.

Experiment 1

In this experiment the degrees of typicality of variants of dimensional values of the artificial stimuli were established. This was done by means of two tasks: a comparison task of the "which of these two variants is more representative of this value?" type, and a direct ranking task. If exemplariness has to do with internal structure, then the comparison task and ranking task had to result in the same ranking orders. It was also hypothesized that differences in difficulty of deciding among the variants would emerge as an (inverse) function of the typicality gap between the two variants, because it is harder to decide among two variants which hardly differ in typicality than among quite dissimilar variants. Hence, an increased difference in typicality of the variants had to result in (1) a decrease in the deci-

sion time (RT) and, (2) an increase in the estimated certainty of the decision (CE). Furthermore, it was supposed that an increase in the difficulty of the decision is accompanied by an increase in information processing activity. Therefore, if GSR primarily indicates uncertainty reduction about alternative hypotheses and to a minor extent information processing activity, no relationship between GSR and the discrepancy in typicality of the variants had to emerge in the comparison task, for in this task no uncertainty reduction about alternatives was involved.

Method

Subjects

Eight male and eight female student volunteers, prescreened for colorblindness, served as Ss. Ss were paid Hfl. 20 for their services.

Stimuli

Four three-valued dimensions were used: figure (triangle, quadrangle, pentagon), letter (alpha, beta, pi), color (blue, brown, green) and arabic numeral (7, 8, 9). Figures were line drawings varying from basic forms to irregular formations. Letters and digits were represented by different handwritings. Variants were obtained in a pilot study by asking Ss to put down on a piece of paper an example of each of the values. Color was varied in rectangles of various shades, which were actually different paint-samples of the three above mentioned colors. From each dimensional value, six variants were chosen which were supposed to differ in typicality. In the comparison task all six variants of each value were compared with each other in pairs. The stimuli consisted of slides on top of which the name of a dimensional value was mounted, and under which one variant was placed on the left and another on the right, labeled A and B respectively. Each variant was put in each position equally often. Within each value, 15 pair comparisons had to be made. In all, there were 12 series of 15 slides, each series being randomized separately.

In the ranking task $3\frac{1}{2} \times 5$ -inch cards were used. Each card represented one variant. There were 72 different cards.

Apparatus

The S was seated in a dimly illuminated, soundproof room. In front of the S was a frosted glass window. Slides were projected onto this window via a carousel projector located outside the room. On the arm rest of the chair a response panel was installed within reach of the S's preferred hand. On this panel several buttons were fixed: a starter button below, two choice buttons, labeled A and B, in the middle, and, at the top, five buttons labeled 1 to 5, representing different degrees of certainty about the choice. Response times (RT) were measured from the beginning of stimulus presentation to S's choice response. RTs were recorded automatically, as were choice responses and certainty ratings.

Basic level GSR was measured DC and specific GSRs were measured AC by a constant 0.5 V voltage bridge. AC responses were registered with a time constant of 3 sec. The greatest conductance change beginning between 1 and 4 sec after the

onset of slide projection was used to calculate the change in log conductance ($\Delta \log C$). $\Delta \log C$ was taken as a measure for GSR to reduce individual differences. Ag/AgCl electrodes, 7.5 mm in diameter, were attached to the volar surface of the distal phalanx of the second and third finger of the *S*'s nonpreferred hand. The electrodes were attached with agar-agar electrode paste with a 0.067 M KCl solution. GSR, choice responses, certainty ratings and projector impulses from the timer were all recorded on a Beckman eight-channel polygraph (type R 411 dynograph).

Design and procedure

Each *S* was given four blocks of three tasks. In each task all variants within a value had to be compared in pairs. For example, the *Ss* judged all 15 possible combinations of the six variants of triangles. A few minutes rest was allowed between the blocks.

The blocks were preceded by a training task in which drawn variants of a tree had to be judged. The order of presentation of the blocks was counterbalanced in four groups. Within each group, two men and two women participated as *Ss*.

After the *S* was seated and GSR electrodes were attached, instructions were given. *Ss* were asked to choose from the two variants the one that was the more representative one of the values indicated by the word on top of the slide. They were told how to give their certainty estimates (CE) of their choice. The task was self-paced: two seconds after the *S* pressed the starter button the slide was projected, and the slide disappeared when the *S* pressed one of the choice buttons. Five seconds after the stimulus disappeared all five certainty buttons lighted up, indicating that the *S* had to give his CE. The lights under these buttons went off when the starter button, which started the next trial, was pressed.

After the *S* completed the four trial blocks, he was given the ranking task. For each dimensional value the six variants had to be ranked according to typicality.

Results

From the pair comparisons a ranking of the variants for each *S* was established. This was done for each dimensional value, by counting the number of times a variant was chosen as the more representative one. To test the consistency of the rankings between *Ss*, the *Ss* were divided into two groups and Spearman rank correlations were calculated on the mean ranks for each value. All correlations turned out to be significant ($r_s \geq 0.87$, $p < 0.03$).

The within subjects consistency was also determined. The ranks from the pair-comparisons were correlated with the rankings in the direct ranking task. All correlations were significant ($r_s \geq 0.83$, $p < 0.04$) except for the color green ($r_s = 0.60$, $p = 0.09$).

For each dependent variable, RT, CE and GSR, four analyses of variance (ANOVA) were performed, one for each dimension. The factors of interest were Order of Presentation, Sex (both between-subject factors), Dimensional Values and Distance between variants (both within-subject factors). Distance was defined as the difference in mean ranks of the variants of each value. Among six variants there are

15 mutual differences. Hence 15 distances could be established, one for each comparison. These 15 distances were divided in a low, medium and high distance category, containing the five low, medium and high distances, respectively.

The ANOVAs did not systematically yield significant effects of Order of Presentation. The same held for the independent variable Sex, except for the GSR analysis for figures, which showed a significantly higher GSR for men than for women ($F(1,8) = 5.92, p < 0.05$).

The ANOVAs showed that the Dimensional Values did not systematically influence RT, CE or GSR. Within each dimension an increasing RT was always accompanied by a decreasing CE, except in one case – the letter dimension. The Distance variable significantly affected RT and CE; for all dimensional values: $F(2,16) \geq 18.92, p \leq 0.001$. RT decreased and CE increased with increasing distance between the variants to be compared. GSR was not influenced by distance, but mean GSR turned out always to be higher at the first presented values than at the latter ones.

Additional analyses were performed to compare the four dimensions. Dimensions significantly affected RT (Friedman two-way ANOVA, $\chi^2 = 8.18, df = 3, p < 0.05$) but not GSR and CE. RT decreased and CE tended to increase from color, to digit, to figure, to letter.

Discussion

A high agreement in ratings was found between Ss as well as within Ss. These results are consistent with the theory of internal structure (Rosch 1973, 1975a). Because of the high consistency in ratings, the stimulus material seems quite useful for experiment 2, in which the typicality of the instances was varied systematically.

The predicted decrease of RT and increase of CE with increasing distance emerged in the data. This indicates that typicality differences are reflected in RT and CE. Apart from distance, the results also showed that RT, and CE in a minor way, were affected by dimensions. This implies that the distances between variants did differ for the various dimensions. However, the relationship between RT and CE was not influenced by these differences in distance: for nearly all dimensional values an increasing RT was accompanied by a decreasing CE.

Distance turned out not to influence GSR in a systematic way, nor did dimensions. However, GSR was influenced by values. The first presented values were associated with higher GSRs than the two latter ones. Habituation of GSR can easily explain this finding.

The results (1) that RT and CE were influenced by distance, whereas GSR was not, and (2) that dimensions affected RT and CE, but not GSR, show that GSR does not vary systematically with task difficulty, where RT and CE do. Another explanation of the lack of relationship between GSR and distance is that the Ss invested an equal amount of effort in the present task each time (Kahneman 1973), irrespective of difference in difficulty of the comparisons to be made, *i.e.* differences in distance. However, this argument fails to explain the RT differences as a function of distance. Hence, the explanation that GSR does not indicate differences in task difficulty seems the most plausible one. In experiment 2 this question will be elaborated further.

Experiment 2

Experiment 2 was designed firstly to investigate the effect of typicality of dimensional values on CI performance, using meaningless stimuli. The following ranking of difficulty of the CI problems was expected: (R typ & I atyp) < (R typ & I typ) < (R atyp & I atyp) < (R atyp & I typ).

The second aim of the study was to elaborate further on the relationship between GSR, uncertainty reduction and information processing activity. The assumption that GSR primarily indicates uncertainty reduction about alternative hypotheses resulted in the prediction (1) that GSR has to vary among different types of feedback: GSR being highest at disconfirmation and lowest at blanks; (2) that GSR does not increase with increasing difficulty of the task, hence no difference in GSR was expected among the four kinds of tasks. If, on the other hand, GSR also indicates the amount of information processing activity involved, then GSR has to be a function of both type of feedback and difficulty of the CI problems.

The final purpose of this study was to test Levine's multi-hypothesis sampling theory. It was predicted that RT would decrease from TLE to ST, and remain constant after ST.

Method

Subjects

Sixteen male students who did not participate in experiment 1, volunteered as Ss. They were prescreened for color-blindness and participated for Hfl. 20 payments.

Stimuli

All variants of the four three-value dimensions described in experiment 1 were used. The six variants of each value were divided into two groups: a group of three variants with the highest representation ratings and a group of three variants rated lowest in experiment 1. These groups are formed by good (typ) and poor (atyp) examples, respectively. The stimulus material consisted of slides. Each slide was divided into four quadrants. The slides showed one variant of each of the four dimensions per quadrant. The variants were placed at random in one of the quadrants. An example is given in fig. 1. The total stimulus population consisted of the 104.976 combinations possible with 6 variants, each of 3 values of 4 dimensions.

Variants of both the relevant (R) and the irrelevant (I) dimensions were either typical or atypical. This resulted in four kinds of tasks: (R typ & I typ), (R typ & I atyp), (R atyp & I typ), and (R atyp & I atyp). In each kind of task four different conjunctive concepts were employed in such a way that two different values of each dimension were used. Hence, there were 16 different tasks. A training task was added, which was composed of typ and atyp variants for both R and I dimensions. Ss received informative (positive or negative) feedback in 67% of the cases, whereas in the remaining cases no informative feedback was presented. The feedback slides showed 'positive', 'negative' or 'unknown'. The composition and the order of

Fig. 1. Example of the material presented to subjects in experiment 2. The dimensions figure, arabic numeral, letter and color are shown, with values pentagon, eight, beta and blue, respectively.

presentation of the stimuli within each task were determined at random with the following restrictions: (1) on average, two values per trial had to change, varying from one to three values at a time; (2) all variants allowed in a task had to be used in about equal numbers; (3) 50% of the stimuli had to be instances of the concept; (4) 'unknown', 'positive' and 'negative' feedback slides had to appear equally often; (5) two orders of feedback were used, both equally often; (6) for an optimal information processor the third trial had to reduce the number of possible hypotheses to one in each task.

Apparatus

The same experimental equipment as in experiment 1 was used. However, the response panel was different. Two choice buttons were used. These were labeled + and -. A lever was fixed at the side of the panel and could be moved forwards and backwards. *Ss* were informed that the range of this lever represented a continuum of uncertainty about their choice response. RT was measured as in the first experiment. This was also true for GSR, except that it was measured as the conductance change beginning between 1 and 4 sec after the onset of feedback instead of stimulus projection.

Design and procedure

All *Ss* were given all four kinds of tasks (R typ & I typ), (R typ & I atyp), (R atyp & I typ), and (R atyp & I atyp), each task with a different concept. Orders of presentation of tasks and concepts were counterbalanced. Within each of four different orders of tasks there were four orders of concepts. Each *S* received a different combination of order of task and concept. The four tasks were preceded by a training task.

After the GSR electrodes had been attached and the *S* was seated, instructions were given. *Ss* were informed of the nature of the task and the means of responding. They were told that a conjunctive concept had to be learned, and that both feedback with information (positive or negative) and feedback without information (unknown) would be given. It was stressed that the actual place of the four values on the slide was irrelevant to the solution. An example of classification of some kinds of fruit was given to indicate the variation within values, which was, however, not stressed in the instruction.

The trials within the tasks started with presentation of the stimulus. The stimulus disappeared when the *S* gave his categorization response by pressing one of the two buttons. If he did not respond within 15 sec the stimulus also disappeared and the *S* had to give his categorization response immediately. Following the categorization response the *S* had to give his certainty rating by pushing the lever to the point that corresponded with his estimation. Four sec after stimulus switch off a little green lamp came on, indicating that the response had to be completed. Again 4 sec later, feedback was shown during 5 sec. The little green light kept burning until feedback switch off. Three sec after feedback switch off the next stimulus was presented.

The criterion for termination of a task was ten successive correct classifications or after 40 trials. At the conclusion of each task, including the training task, the *S* was shown four boards. On each of them all cards of one dimension of the card sorting task from experiment 1 were arranged randomly. These cards represented all possible variants (typ and atyp). The *S* was asked to compose a representative example of the concept just learned from these cards.

Results

Trials to the last error (TLE) and number of errors (NE)

Results for TLE and NE are shown in table 1. Overall mean TLE and NE were 17.02 and 6.19, respectively. ANOVAs of both TLE and NE showed a significant effect of Tasks ($F(3,36) = 7.17$ and 7.18 , respectively, both $p < 0.001$). *A posteriori* comparisons showed: (R typ & I atyp) = (R typ & I typ) < (R atyp & I atyp) = (R atyp & I typ) for TLE as well as for NE (Duncan's new multiple range test, $p < 0.05$). Furthermore, comparison of both I typ conditions with both I atyp conditions revealed that both TLE and NE for these conditions did not differ significantly from each other, although TLE showed a tendency to be higher in the I typ than in the I atyp conditions (sign test, $p < 0.06$).

No effect of Order of Presentation on the tasks emerged; the interaction of Order of Presentation $\times$ Tasks was not significant ($F(9,36) = 1.06$ and 0.89 for TLE and NE, respectively).

The number of *Ss* who did not solve the problem within 40 trials was 3, 1, 5 and 4 in the conditions (R typ & I typ), (R typ & I atyp), (R atyp & I typ), and (R atyp & I atyp), respectively. These numbers did not differ significantly (Cochran Q test; $Q = 3.62$, $df = 3$).

Table 1
Mean values of the main effects for each of the dependent variables of experiment 2

Main sources	Dependent variables				
	Trial of last error	Number of errors	Galvanic skin response (in $\Delta \log C \times 100$)	Reaction time (in sec)	Certainty estimate
Tasks					
A (R typ + I typ)	12.88	4.69	7.35	5.51	0.61
B (R typ + I atyp)	6.79	2.75	7.21	5.78	0.55
C (R typ + I typ)	25.19	8.69	7.98	5.98	0.54
D (R typ + I atyp)	23.31	8.63	7.96	5.71	0.59
Order of presentation					
A B C D	17.31	7.13	8.15	5.86	0.55
B A D C	18.25	6.44	5.38	5.18	0.48
C D A B	15.69	5.63	7.40	6.11	0.61
D C B A	16.81	5.56	9.58	5.83	0.64
Type of feedback					
Blank	—	—	6.19	—	—
Confirmation	—	—	7.78	—	—
Disconfirmation	—	—	8.91	—	—
Overall mean	17.02	6.19	7.63	5.74	0.57

GSR

The ANOVA of GSR data showed no significant difference between the four tasks ($F(3,36) = 0.44$). However, there was a significant difference between GSR at blanks, confirming feedback and disconfirming feedback ($F(2,24) = 4.49$, $p < 0.02$). GSR data are shown in table 1. GSR was based on the same number of trials for the three types of feedback per S per task. This was done to reduce possible differences caused by frequency or habituation. *A posteriori* comparisons showed that GSR after blanks tended to be lower than after confirmations ($p < 0.10$), and was lower than after disconfirmations ($p < 0.01$); no significant difference between GSR after confirmations and after disconfirmations emerged (Duncan's new multiple range test).

The Order of Presentations $\times$ Tasks interaction was significant ($F(9,36) = 2.11$, $p < 0.05$), which may be considered as an effect of habituation between tasks, because learning a task later on resulted in a smaller GSR. The effect of habituation within tasks was investigated by Vincentizing each of the tasks in equal parts for each S to correct for the difference in number of trials between S s. No significant

effect of habituation within each of the four tasks emerged (Friedman two-way ANOVA, $\chi^2 \leq 3.22$, $df = 2$).

To determine the relationship between the number of confirmations directly preceding a disconfirmation and GSR at that disconfirmation, the GSR data were classified into four categories according to the number of preceding confirmations (0, 1, 2 and ≥ 3). These categories were about equally large. The number of confirmations was based on confirmations apart from blanks between the confirmations. GSR did not increase with an increasing number of confirmations (Friedman two-way ANOVA, $\chi^2 = 1.73$, $df = 3$). Nor was this increase significant when blanks were included. Because of a lack of data the GSR increase could not be tested for confirmations not interrupted by blanks.

Investigation of the relationship between CE and GSR yielded the following results. GSR at feedback on trial n did not change when certainty either increased or decreased at trial $n + 1$ (Friedman two-way ANOVA, both $\chi^2 = 1.63$, $df = 2$). However, there was a tendency to increasing GSR with increasing discrepancy between CE and actual feedback. This discrepancy was measured at six points: confirmation of a categorization response with high, medium and low CE and disconfirmation of a response with low, medium and high CE. GSR tended to be positively related to this order of increasing discrepancy (Friedman two-way ANOVA, $\chi^2 = 10.60$, $df = 5$, $p < 0.10$).

Reaction time and certainty estimate

Results for RT and CE are shown in table 1. Overall mean RT and CE were 5.74 and 0.57 respectively. In the ANOVAs, no significant effect of Tasks was found ($F(3,36) = 0.36$ and 1.22 , respectively). The Order of Presentation $\times$ Tasks interaction was not significant ($F(9,36) = 1.99$ and 1.96 , respectively, both $p < 0.07$) although tasks presented later tended to show faster RTs and higher CEs.

To test Levine's multiple-hypothesis sampling theory, RT data before and after TLE as well as before and after the solution trial (ST) were analyzed. The first trial following TLE with a 100% CE was defined as the ST. No difference in RT was obtained in the last 5 trials up to and including TLE (Friedman two-way ANOVA, $\chi^2 = 2.42$, $df = 4$). However, RT decreased significantly from TLE up to and including the fourth trial following TLE ($\chi^2 = 24.53$, $df = 4$, $p < 0.001$). RT also decreased significantly in the last 5 trials up to and including ST, but after TLE ($\chi^2 = 14.97$, $df = 4$, $p < 0.01$). Finally, no difference was found in the trials from ST up to and including the fourth trial following ST ($\chi^2 = 8.05$, $df = 4$, $p < 0.10$), although RT tended to decrease.

Representative example

In the CI task Ss were presented with either typ or atyp variants of a particular value. Directly after completing each CI task Ss were asked to give a representative example of a positive instance. Ss significantly more frequently gave examples containing typ variants of values that were actually represented only by atyp variants during the CI task, than they gave atyp variants as examples if they had seen only typ variants of those values (McNemar test for the significance of changes, $\chi^2 =$

5.82, $df = 1$, $p < 0.02$). The former occurred in 31% of the cases, the latter in 3% of the cases.

Discussion

The results demonstrated effect of qualitative intra-value variation on CI performance. The degree of typicality of dimensional values influenced performance in a significant way. Typical values led to faster CI than atypical values, because (R typ & I typ) was easier to solve than (R atyp & I atyp). However, the effect appeared to be limited to relevant values. Typicality of relevant values facilitated CI, whereas for irrelevant values performance was not affected by typicality. An explanation for this might be that the relevant values are more open to a typicality effect because correct solution requires a more exact distinction between variants of relevant values than between irrelevant ones. The influence of typicality on CI performance corresponds with Rosch's (1973) theory of internal structure. The same holds for the findings of the additional task. After CI, Ss frequently chose typical variants as most representative values of the concept if they had seen only atypical variants of those values. They did so more often than they chose atypical variants if, during CI, they were actually shown only typical variants. In other words, Ss identified the concept more in typical than in atypical variants, even if typical variants had not actually been presented to them in the learning task.

The second topic of this study was the relationship between the amount of information processing and GSR. GSR varied among different types of feedback; GSR was highest at disconfirmation and lowest at blanks. This finding is consistent with the earlier findings of De Swart and Das-Smaal (1976, 1979a,b). However, in the present experiment, the difference in GSR between confirming and disconfirming feedback failed to reach significance, although the difference was in the predicted direction. Furthermore, GSR tended to increase with increasing discrepancy between expected and actual feedback, as measured by certainty ratings and type of feedback. Such a relationship was found before by De Swart and Das-Smaal (1979b). Squires *et al.* (1973) and De Swart *et al.* (in preparation) found this relationship for an EEG response, the P3, at a perceptual decision task and in a CI task, respectively. The results are in agreement with Sokolov's (1969) model of the OR. The tendency of GSR to increase with increasing discrepancy between expected and actual feedback can also explain why a direct relationship between GSR and CE did not emerge. If confirming and disconfirming feedback are taken together, both the largest and the smallest GSRs are elicited by feedback following the highest certainty.

GSR did not rise with increasing difficulty of the CI tasks. The above mentioned relationship between GSR and type of feedback, and the lack of variation of GSR with task difficulty, together with the lack of relationship between GSR and the discrepancy in typicality in the comparison task of experiment 1, justify the conclusion that GSR indicates uncertainty reduction about alternative hypotheses rather than information processing activities *per se*.

GSR at disconfirmation did not vary with the number of preceding confirmations. De Swart and Das-Smaal (1979a) found the same result in the condition of 67% feedback, which applied also to the present experiment. However, in the 100%

feedback condition of the former experiment a positive relationship was found between number of preceding confirmations and GSR at disconfirming feedback. Another study by De Swart and Das-Smaal (1976), in which 100% feedback was also given, yielded the same result. As an explanation of the difference between the 100% and 67% feedback conditions, De Swart and Das-Smaal (1979a) suggested an inclination to lapse into a passive state after a blank, as was done by Falmagne (1970) and Aiken *et al.* (1972). In this state the *S* is more apt to respond by chance, while this has no effect on the learning process *per se*. In the 100% feedback condition the preceding confirmations are not interrupted by blanks. In this situation the relationship between number of confirmations and GSR at disconfirmation did hold. This, again, supports the idea that GSR primarily reflects uncertainty reduction.

Finally, the present study revealed a decrease in RT from TLE to ST, whereas in the five trials before TLE, RT remained constant. In the trials following ST, RT decreased, though not significantly. Levine's multi-hypothesis sampling theory predicts a decrease in RT between TLE and ST, but no decrease after ST, because then only one hypothesis is left. Except for the slight decrease in RT following ST, the present results support Levine's theory, that subjects monitor a subset of hypotheses, which is narrowed down to one, the correct hypothesis. The tendency of RT to decrease following ST may be explained by an increase of the ease with the *S* applies his hypothesis about the correct concept to new stimuli, which in turn influences decision time.

Experiment 3

The purpose of this experiment was to determine degrees of typicality for variants of dimensional values of real-life stimuli. Experiment 1 showed a high agreement between the paired comparisons and the ranking task. Therefore, it was decided to use only the more easily executable ranking task.

Method

Subjects

Ten female and ten male psychology freshman, who had not participated in the former experiments, served as *Ss*. Participation in the experiment was considered part of their training.

Stimuli

Three three-valued flower dimensions were used: shape of blossom (separate petals, cup-shaped, fluffy with dense pubescence), shape of leaf (rounded, pointed, notched) and stem (smooth, hairy, thorny). From each dimensional value nine variants were chosen which were supposed to differ in typicality. All of them were copies from Nordens Flora (Lindman 1964). Blossoms were colored yellow, leaves and stems were colored green and brownish green, respectively. Each variant was represented on a $3\frac{1}{2} \times 5$ inch card. In all there were 3 decks of 27 cards, one deck for each dimension.

Design and procedure

Each *S* had to divide each deck of cards into three equal groups of cards belonging together in his opinion. After division of a deck the *S* was asked to describe each of the three groups. The *S* then had to rank each group separately according to typicality. Each deck was shuffled before it was given to the *S*. The order in which the decks had to be treated was counterbalanced in four groups. Within each group about an equal number of men and women served as *S*s. Four different orders were given because a fourth dimension, color, was included. This dimension was at first intended to color the blossoms in the next experiment. However, it was decided to keep the color of the blossom constant. Therefore, the color data are not reported.

Results

From the groups into which *S*s divided the decks of cards and from the descriptions they gave of these groups, it was inferred that categorization had taken place according to the dimensional values described above. However, a few times some variants were confused. Variants that were assigned by more than one of the *S*s to values of which they were not intended to be variants, were excluded from the next experiment. These were 12 of the 81 variants: 4 blossom, 3 leaf and 5 stem variants. From the remaining variants 6 of each value were selected for use in the next experiment. These were the three variants of each value rated, on average, highest and the three rated lowest in typicality. To test the consistency of rankings of the 6 selected variants between *S*s, the *S*s were divided into two groups and Spearman rank correlations were calculated in the mean ranks for each value. All correlations were significant ($r_s > 0.76$, $p < 0.05$), except for the smooth value of the stem dimension ($r_s = 0.54$, $p = 0.11$).

Discussion

A good agreement in ratings was found between *S*s for almost all values. As in experiment 1, these results are consistent with Rosch's (1973, 1975a) theory of internal structure. The stimulus material selected as mentioned above was considered suitable for the purpose of experiment 4.

Experiment 4

Experiment 4 was designed to investigate whether stimulus material influences concept identification performance. Therefore, the experimental structure of experiment 2 was replicated with different stimulus material; more realistic and unitary stimuli were employed instead of stimuli composed of artificial, unconnected, separable dimensions.

Method

Subjects

*S*s were 16 male psychology freshmen who had not participated in the previous

Fig. 2. Example of the material presented to subjects in experiment 4. The dimensions shape of blossom, shape of leaf, stem and color of flower-pot are shown, with values separate petals, pointed, thorny and brown, respectively.

experiments. Participation was considered part of their training. Ss were pre-screened for color-blindness.

Stimuli

Stimuli were composed of the variants of four three-value dimensions. The three dimensions tested in experiment 3 were used, together with the color dimension tested in experiment 1. The latter one was used to color a flower-pot. Dimensional values were blossom with separate petals, cup-shaped and fluff-shaped blossom, rounded, pointed and notched leaf, smooth, hairy and thorny stem, and blue, brown and green colored flower-pot. Variants of these values of each dimension were joined and mounted on slides. Each slide showed one blossom with one leaf and stem in a flower-pot. An example is given in fig. 2. Each dimensional value had six variants, three variants with the highest representation ratings and three variants rated lowest in the former experiments. These groups can be considered as good

(typ) and poor (atyp) examples, respectively. Variants of both the relevant (R) and the irrelevant (I) dimensions were either typical or atypical. This resulted in four kinds of tasks: (R typ & I typ), (R typ & I atyp), (R atyp & I typ), and (R atyp & I atyp). The composition of the stimuli and tasks, and the order of presentation was the same as in experiment 2, except that, because of the different dimensions, the place of the variants was constant instead of randomized.

Apparatus

The same experimental equipment as in experiment 2 was used.

Design and procedure

The same design and procedure as in experiment 2 was used. However, the instruction was adapted to the new stimulus material.

Results

Trials to the last error (TLE) and number of errors (NE)

Results for TLE and NE are shown in table 2. Overall mean TLE and NE were 21.88 and 7.47, respectively. A significant effect of Tasks turned up in the ANOVAs of both TLE and NE ($F(3,36) = 5.53$ and 5.67 , respectively, both $p < 0.001$). *A posteriori* comparisons showed for TLE and NE: (R typ & I typ) and (R typ & I atyp) $<$ (R atyp & I typ) ($p < 0.01$, respectively), and furthermore fore TLE (R typ & I typ) $<$ (R atyp & I atyp) ($p < 0.05$); the other pair comparisons between the four tasks did not reach significance (Duncan's new multiple range test). Comparisons of both R typ together, with both R atyp conditions together revealed a lower NE for R typ than for R atyp conditions (sign test, $p < 0.02$); for TLE a tendency to this difference was found (sign test, $p < 0.06$). Comparisons of both I typ with both I atyp conditions showed no significant differences (sign test).

No effect of Order of Presentation on the tasks was found; the Order of Presentations $\times$ Tasks interaction was not significant ($F(9,36) = 0.82$ and 1.10 for TLE and NE, respectively).

The number of *Ss* who did not solve the problem within 40 trials was 2, 3, 9 and 6 in the conditions (R typ & I typ), (R typ & I atyp), (R atyp & I typ), and (R atyp & I atyp), respectively. These numbers differ significantly (Cochran Q test; $Q = 8.57$, $df = 3$, $p < 0.05$). Differences in TLE and NE between experiment 2 and 4 appeared to be not significant (Mann-Whitney U Test, $U = 83.5$ and 96.5 , respectively).

GSR

GSR data for the ANOVA were based on the same number of trials for the three types of feedback per *S* per task. This was done to reduce possible differences caused by frequency or habituation. The ANOVA revealed no significant differences in GSR between the four tasks ($F(3,36) = 0.67$). However, a significant difference was found between GSR at blanks and confirming and disconfirming feedback ($F(2,24) = 13.08$, $p < 0.001$). Mean GSRs are shown in table 2. GSR increased

Table 2

Mean values of the main effects for each of the dependent variables of experiment 4

Main sources	Dependent variables				
	Trial of last error	Number of errors	Galvanic skin response (in $\Delta \log C \times 100$)	Reaction time (in sec)	Certainty estimate
Tasks					
A (R typ + I typ)	13.44	4.88	7.15	6.50	0.54
B (R typ + I atyp)	17.38	5.44	6.88	6.24	0.52
C (R atyp + I typ)	31.56	11.25	8.21	5.70	0.56
D (R atyp + I atyp)	25.31	8.31	6.29	6.34	0.51
Order of presentation					
A B C D	15.81	4.06	6.71	6.25	0.56
B A D C	26.19	8.88	8.67	6.47	0.53
C D A B	19.25	7.69	4.79	6.36	0.43
D C B A	26.25	9.25	8.35	5.70	0.61
Type of feedback					
Blank	—	—	5.75	—	—
Confirmation	—	—	7.30	—	—
Disconfirmation	—	—	8.34	—	—
Overall mean	21.88	7.47	7.13	6.19	0.53

from blanks to confirming to disconfirming feedback. *A posteriori* comparisons showed a lower GSR after blanks than after confirmations ($p < 0.01$) and after disconfirmations ($p < 0.002$), and a lower GSR after confirmations than after disconfirmations ($p < 0.05$) (Duncan's new multiple range test).

The Order of Presentation $\times$ Tasks interaction was not significant ($F(9,36) = 1.45$). Habituation within tasks was investigated by Vincentizing each of the tasks in equal parts for each *S* to correct for the difference in number of trials between *Ss*. No significant effect of habituation within each of the four tasks emerged (Friedman two-way ANOVA, $\chi^2 \leq 1.53$, $df = 2$).

To determine the relationship between the number of previous confirmations and GSR at the next disconfirmation, the GSR data were classified into four categories according to the number of confirmations directly preceding a disconfirmation (0, 1, 2 and ≥ 3). These categories were about equally large. Number of confirmations did not include the intervening blanks. GSR did not increase with increasing number of confirmations before disconfirmation (Friedman two-way ANOVA, $\chi^2 = 3.51$, $df = 3$). Nor was this relationship significant when blanks were included in the number of confirmations. The relationship also failed to reach

significance when only confirmations not interrupted by blanks were considered ($\chi^2 = 1.78, df = 2$).

The following results were found for the relationship between GSR and CE. GSR at feedback on trial n did not change when certainty either increased or decreased at trial $n + 1$ (Friedman two-way ANOVA, $\chi^2 = 3.50$ and 1.13 , respectively, $df = 2$). Also no significant difference in GSR was found as a function of the discrepancy between the degree of certainty of the S and the actual feedback, measured at six points as described in experiment 2 (Friedman two-way ANOVA, $\chi^2 = 5.26, df = 5$).

Reaction time and certainty estimate

Results are shown in table 2. Overall mean RT and CE were 6.19 sec and 0.53, respectively.

In the ANOVAs, neither a significant effect of Tasks ($F(3,36) = 2.09$ and 0.46 , respectively), nor a significant effect of Order of Presentation on the different tasks as measured by the Order of Presentation $\times$ Tasks interaction ($F(9,36) = 1.77$ and 1.10 , respectively) was found.

To test Levine's multiple-hypothesis sampling theory, RT data before and after TLE as well as before and after ST were analyzed in the same way as described in experiment 2. No differences in RT were obtained in the 5 trials before TLE, nor in the 5 trials after ST (Friedman two-way ANOVA, $\chi^2 = 3.32$ and 6.25 , respectively, $df = 4$). However, RT decreased significantly after TLE and before ST ($\chi^2 = 22.06$ and $11.26, df = 4, p < 0.001$ and $p < 0.025$, respectively).

Representative example

In the additional task S s were asked to construct a representative example of the concept just learned. In 41% of the cases, these examples contained typical variants of values which were represented in the CI task only by atypical variants. This tended to occur more frequently than the opposite, examples containing atypical variants of values of which only typical variants had been shown, which occurred in 16% of the cases. (McNemar test for the significance of changes, $\chi^2 = 2.72, df = 1, p < 0.10$). In all, the examples given by S s in experiment 2, significantly less frequently contained variants which actually had not been shown during the CI task than they did in the present experiment ($\chi^2 = 4.45, df = 1, p < 0.05$).

Discussion

As in experiment 2, the degree of typicality of dimensional values influenced performance. Typical values resulted in faster CI than atypical values. Furthermore, typicality of relevant values tended to facilitate CI, but again, variation of typicality of irrelevant values did not influence performance. The additional task revealed that in the present experiment S s tended to choose more typical than atypical variants as representative examples of the concept. This finding also corresponds with Rosch's concept of internal structure of categories, where typical variants form the focus of organization.

The second prediction of the experiment concerned the relationship between GSR and information processing. Again, the results contributed to support the idea that GSR primarily indicates uncertainty reduction. GSR was not influenced by difficulty of the tasks. However, GSR did vary with type of feedback, being highest at disconfirmation and lowest at blanks. Only the discrepancy between expected and actual feedback failed to reach significance in the present experiment.

Like the earlier findings in 67% feedback situations, this experiment showed no relationship between GSR at disconfirmation and number of preceding confirmations. As suggested in the discussion of experiment 2, a passive state after a blank offers an explanation for this result.

Levine's multi-hypothesis sampling hypothesis was tested by investigation of the RT data before and after TLE as well as before and after ST. RT remained constant before TLE and after ST, but RT decreased from TLE to ST. This result supports Levine's theory about the sampling of hypotheses.

General discussion

The principle of internal structure, as suggested by Rosch (1973), states that not every instance of a category is as good an exemplar as any other one. Category members differ in typicality, and this has consequences for information processing. Rosch (1973) stated that applicability of the concept of internal structure to categories refers to two issues: "can subjects make meaningful judgments about internal structure — that is, about the degree to which instances are "focal" members of categories; and can a reasonable case be made that internal structure affects cognitive processing of the categories?" (1973: 130). Both conditions are satisfied by the results of the present four experiments. The first question is answered by the results of experiments 1 and 3. In these experiments a high agreement in judgment was found. Furthermore, in experiment 1, the distance between different degrees of typicality appeared to be reflected in RT and CE. The second demand is fulfilled by the results of experiments 2 and 4. In these experiments it was shown that typicality of dimensional values influenced CI performance. In all, the results are clearly in accordance with the concept of internal structure.

A remark has to be made about the atypical instances actually used. Rosch's theory states that atypical instances are the more peripheral ones. These instances have more in common with contrasting categories. The more a variant of a particular value shares with another category, the higher the chance that it will be classified as one of that category.

Classification of an atypical variant in a contrasting category would change a conjunctive CI task into some kind of a disjunctive one. This would actually have made those tasks unsolvable, in which the relevant values were represented by atypical variants, because the subjects were told to look for a conjunctive solution. This possibility was ruled out by the experimental set up of experiment 2 and 4, and is not evidenced by the data of experiment 2. In experiment 1 and 2 only those atypical variants of a value were chosen, which were thought to be no examples of other values. The data of experiment 2 showed, that concepts learned by atypical variants of relevant values were solved equally often as those learned by typical variants. The stimulus material used in experiment 3 and 4 initially seemed to be somewhat more susceptible to misclassification. In experiment 3, variants had to be classified according to value. The variants of a given value which were not classified as a variant of that value by more than one subject, were not used in experiment 4. Hence, it is unlikely that the results obtained could be ascribed to an artifact of the stimulus material.

The way in which the prototypes are formed did not influence the consistency of rankings in experiment 1. In this experiment the dimensions used were color, form, letter and digit. According to Rosch and Mervis (1975) the color and form dimensions are categories which probably have a physiological basis, with prototypes formed by stimuli which are salient prior to the formation of the category. On the other hand, prototypes of the other two dimensions, letter and digit in different handwritings, must be formed through principles of learning. Experiment 1 showed that the concept of internal structure may apply also to learned dimensions.

According to Rosch (1978), prototypes themselves only constrain but do not specify representation and process models. In other words, models should not be inconsistent with the known facts about typicality. The results of the present study provide counterevidence against any CI model that treats instances of a category as equivalent in their degree of membership of the category.

The second topic of the present study concerned the GSR. The question whether GSR indicates primarily uncertainty reduction or also information processing activities was clarified by the present results. GSR did not vary systematically in experiment 1, which involved different amounts of information processing activities but no decisions about alternative hypotheses. GSR in experiments 2 and 4 did vary with

amount of uncertainty reduction, but not with task difficulty. Together, these results favor the uncertainty reduction hypothesis. An explanation of the results in terms of habituation can be simply refuted by the lack of habituation within the tasks of experiments 2 and 4. Frequency of occurrence of the three types of feedback also offers no explanation, because GSR was higher at confirmation than at blanks, whereas blanks occurred less often than confirmations. Besides, a correction was made to reduce possible differences caused by frequency or habituation by using the same number of trials for the three types of feedback per subject per task. The uncertainty hypothesis remains the best explanation of the results.

De Swart and Das-Smaal (1979b) found different relationships between uncertainty reduction and GSR after each type of feedback. When confirming feedback followed a classification, GSR was larger when the subject was uncertain about the classification than when he was certain. After disconfirming feedback, GSR showed the opposite tendency. After non-informative feedback, GSR was not influenced by uncertainty. In the present experiment these relationships did not emerge. An explanation for the difference in results might be that in the present study certainty ratings were not determined by certainty about hypotheses alone, but also by typicality; low typicality could have resulted in lower CE than high typicality.

Influence of stimulus material on CI performance did not show up in overall mean TLE or NE; neither TLE nor NE differed between experiment 2 and 4. Furthermore, in both experiments CI performance was better with typical values than with atypical values. A difference between both experiments turned up in the additional task. The representative example given by subjects after CI, contained variants which actually had not been shown in the CI task more often for the realistic stimuli than for the artificial stimuli. This result could have been caused by smaller differences in typicality between variants in experiment 4 than in experiment 2.

Finally, both in experiment 2 and 4, support was found for Levine's multihypothesis sampling theory. This theory states that the subject monitors a subset of hypotheses, which is reduced to the only correct hypothesis by rejecting those hypotheses that are disconfirmed by the feedback. The decrease in RT following TLE is predicted by the narrowing down of the subset, but may also be explained in terms of increasing confidence in the working hypothesis. Increasing confidence

in the working hypothesis implies decreasing confidence in other hypotheses, and this goes on until the solution trial, at which by definition, the confidence in the working hypothesis is maximal.

References

- Aiken, L.S., J.L. Santa and A.B. Ruskin, 1972. Nonreinforced trials in concept identification: presolution statistics and local consistency. *Journal of Experimental Psychology* 93, 100–104.
- Archer, E.J., 1962. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology* 63, 616–620.
- Bruner, J.S., J.J. Goodnow and G.A. Austin, 1956. *A study of thinking*. New York: Wiley.
- Cutting, J.E. and N.J. Schatz, 1976. On the relationship between intercategory and intracategory semantic structure. *Bulletin of the Psychonomic Society* 8, 406–408.
- De Swart, J.H. and E.A. Das-Smaal, 1976. Relationship between SCR, heart rate and information processing. *Biological Psychology* 4, 41–49.
- De Swart, J.H. and E.A. Das-Smaal, 1979a. Physiological activity in an attribute learning task. *Acta Psychologica* 43, 57–69.
- De Swart, J.H. and E.A. Das-Smaal, 1979b. 'Orienting reflex and uncertainty reduction in a concept learning task'. In: H.D. Kimmel, E.H. van Olst and J.F. Orlebeke (eds.), *The orienting reflex in humans*. Hillsdale, NJ: Lawrence Erlbaum.
- De Swart, J.H., A. Kok and E.A. Das-Smaal, in preparation. P300 and uncertainty reduction in a concept learning task.
- Falman, R., 1970. Construction of a hypothesis model for concept identification. *Journal of Mathematical Psychology* 7, 60–96.
- Garner, W.R., 1976. Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology* 8, 98–123.
- Goldman, D. and D. Homa, 1977. Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology, Human Learning and Memory* 3, 375–385.
- Kahneman, D., 1973. *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Levine, M., 1969. Latency-choice discrepancy in concept learning. *Journal of Experimental Psychology* 82, 1–3.
- Lindman, C.A.M., 1964. *Nordens flora*. Stockholm: Wahlström and Widstand.
- Posner, M.I., 1964. Information reduction in the analysis of sequential tasks. *Psychological Review* 71, 491–504.
- Rosch, E., 1973. 'On the internal structure of perceptual and semantic categories'. In: T.E. Moore (ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.
- Rosch, E., 1975a. Cognitive representation of semantic categories. *Journal of Experimental Psychology, General* 104, 192–233.
- Rosch, E., 1975b. Cognitive reference points. *Cognitive Psychology* 7, 532–547.
- Rosch, E., 1978. 'Principles of categorization'. In: E. Rosch and B.B. Lloyd (eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum.
- Rosch, E. and C.B. Mervis, 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 7, 573–605.
- Smith, E.E., E.J. Shoben and L.J. Rips, 1974. Structure and process in semantic memory: a featural model for semantic decision. *Psychological Review* 81, 214–241.

- Sokolov, E.N., 1969. 'The modeling properties of the nervous system'. In: M. Cole and I. Maltzman, A handbook of contemporary Soviet psychology. New York: Basic Books.
- Squires, K.C., S.A. Hillyard and P.H. Lindsay, 1973. Cortical potentials evoked by confirming and disconfirming feedback following an auditory discrimination. *Perception and Psychophysics* 13, 25–31.