

TNO Defence Research

**AD-A256 553**



TNO-report **IZF 1992 B-4**

**J.M.C. Schraagen**

TNO Institute for Perception

Kampweg 5  
P.O. Box 23  
3769 ZG Soesterberg  
The Netherlands

Fax +31 3463 5 39 77  
Telephone +31 3463 5 62 11



**REPRESENTATION OF RESEARCH PARADIGMS AS A FUNCTION OF FAMILIARITY WITH RESEARCH DOMAINS**

**DTIC**  
**ELECTE**  
**OCT 28 1992**  
**S A D**

This document has been approved  
for public release and sale; its  
distribution is unlimited.

All rights reserved.

No part of this publication may be  
reproduced and/or published by print,  
photoprint, microfilm or any other means  
without the previous written consent of  
TNO.

In case this report was drafted on  
instructions, the rights and obligations of  
contracting parties are subject to either the  
'Standard Conditions for Research  
Instructions given to TNO', or the relevant  
agreement concluded between the  
contracting parties.  
Submitting the report for inspection to  
parties who have a direct interest is  
permitted.

© TNO

Number of pages: 31

**92-28264**

## CONTENTS

	Page
SUMMARY	5
SAMENVATTING	6
1 INTRODUCTION	7
2 METHOD	11
2.1 Subjects	11
2.2 Stimuli	11
3 RESULTS	13
3.1 Effects of problem familiarity	15
3.2 Effects of understandability	18
3.3 Classification of features	19
4 DISCUSSION	20
REFERENCES	25
APPENDIX A Articles selected, research areas, number of citations	26
APPENDIX B Sentences used as stimuli	29
APPENDIX C Literal instructions for feature listing and sentence classification	31

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

Report No.: IZF 1992 B-4  
Title: Representation of research paradigms as a function  
of familiarity with research domains  
Author: Drs. J.M.C. Schraagen  
Institute: TNO Institute for Perception  
Group: Information Processing  
Date: July 1992  
DO Assignment No.: B92-11  
No. in Program of Work: 787.1

---

## SUMMARY

Previous research (Schraagen, 1990) has shown that experienced researchers use paradigms when designing experiments. The purpose of the present study was to determine how the content or quality of the representation of paradigms improves as a function of problem familiarity. The present study systematically varied problem familiarity for each subject separately, based on the subject's self-reported familiarity with various research domains. Five research domains were chosen for each subject and from each domain a journal article was chosen. One sentence describing the question to be answered in the article was extracted from each of the five articles and presented to the subject. The methods of object categorization and feature listing, originally used in categorization experiments by Rosch and associates (e.g., Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976), were used here to determine the content of subjects' paradigms. Choice of the correct paradigm was assessed by asking the subjects to classify a particular research question as being of a certain type. The content of the paradigm was assessed by asking subjects to write down as many characteristics for this type of research as they could. Thirty-four subjects participated in the experiment.

The results showed that as subjects were more familiar with a particular research area, they used more specific words to classify the area and they listed more features overall. Only when subjects were highly familiar with a research area did they list highly specific features, dealing with how to measure variables, what number and type of subjects to select, what control variables to use, what hypotheses to test, and what possible outcomes to expect. When confronted with problems only slightly outside their area of expertise, experts must rely upon general design knowledge and general knowledge about what are relevant features for the novel area. The present study has also shown that these types of knowledge, and domain knowledge as well, are acquired rather soon after one has specialized in a particular area, given that no differences were found between subjects with three years of experience and subjects with thirty years of experience.

**Representatie van onderzoeksparadigma's als functie van bekendheid met onderzoek**

J.M.C. Schraagen

**SAMENVATTING**

Vorig onderzoek (Schraagen, 1990) heeft laten zien dat ervaren onderzoekers paradigma's gebruiken bij het opzetten van onderzoek. Het doel van het huidige onderzoek was te bepalen hoe de inhoud of kwaliteit van de representatie van paradigma's verbetert als functie van bekendheid met het onderzoek. In het huidige onderzoek werd bekendheid met het onderzoek voor iedere proefpersoon afzonderlijk gevarieerd, gebaseerd op een zelf-rapportage van de proefpersoon. Voor iedere proefpersoon werden vijf onderzoeksdomeinen gekozen en uit ieder domein werd een tijdschrift artikel gekozen. Uit ieder van de vijf artikelen werd één zin gekozen waarin de in het artikel te beantwoorden vraagstelling werd beschreven. Deze vijf zinnen werden aan de proefpersoon voorgelegd. De methodes van object categorisatie en noemen van kenmerken, zoals oorspronkelijk gebruikt in categorisatie experimenten door Rosch en medewerkers (b.v. Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976), werden in dit onderzoek gebruikt om de inhoud van de paradigma's te kunnen vaststellen. Keuze van het juiste paradigma werd bepaald door proefpersonen te vragen de betreffende zin te classificeren als afkomstig van een bepaald type onderzoek. De inhoud van het paradigma werd bepaald door proefpersonen te vragen zoveel kenmerken voor het betreffende onderzoek te noteren als zij konden. Aan het onderzoek deden 34 proefpersonen mee.

De resultaten lieten zien dat als proefpersonen bekender waren met een bepaald onderzoeksgebied, zij meer specifieke woorden gebruikten om het gebied te classificeren en meer kenmerken noteerden. Slechts wanneer proefpersonen zeer bekend waren met een bepaald onderzoeksgebied, noteerden zij zeer specifieke kenmerken. Deze kenmerken behelsden het meten van variabelen, het controleren van storende variabelen, het selecteren van aantal en soort proefpersonen, welke hypothesen getoetst dienen te worden en welke resultaten verwacht mogen worden. Wanneer experts geconfronteerd worden met problemen die slechts weinig buiten het eigen gebied van expertise liggen, vallen zij terug op algemene kennis omtrent het opzetten van onderzoek en algemene kennis over wat relevante kenmerken zijn voor het relatief onbekende onderzoeksterrein. Het huidige onderzoek heeft ook laten zien dat deze soorten kennis, en domeinkennis eveneens, relatief snel verworven worden nadat men zich in een bepaald gebied heeft gespecialiseerd, gegeven het resultaat dat geen verschillen werden gevonden tussen proefpersonen met drie jaar ervaring en proefpersonen met dertig jaar ervaring in het opzetten van onderzoek.

## 1 INTRODUCTION

Previous research (Schraagen, 1990) has shown that experienced researchers use paradigms when designing experiments. Paradigms may be viewed as plans or templates that contain cohesive pieces of design knowledge to be used in particular types of experiments. This design knowledge specifies what subjects, independent variable, dependent variable, and control variables should be used in, for instance, a typical selective attention experiment. Another important finding was that a general strategy such as problem decomposition can be applied by experienced researchers even when they are unfamiliar with the research question. Protocol studies indicated that experienced researchers, when faced with unfamiliar problems, classified these problems as being solvable with a particular kind of paradigm. Interestingly, because of their unfamiliarity with the problem, these researchers often chose the wrong kind of paradigm, as assessed by domain experts. However, choosing the wrong kind of paradigm did not prevent these researchers from applying problem decomposition and maintaining a structured approach to problem solving. Hence, a distinction can be made between the form and the content of reasoning: the form of reasoning may generalize across problems, whereas the content may deteriorate as problems become less familiar.

Although the findings mentioned above were established by objective procedures and could be assessed quantitatively, these findings still are of limited generalizability and limited power, for several reasons. One is that few subjects were used, so that the results may not apply to different samples of experts. A second reason is that familiarity with the domain in which an experiment had to be designed was manipulated between, and not within, subjects. Hence, the results could in principle be attributed as much to the experts' idiosyncrasies as to their domain familiarity. Third, only one problem was used, which may limit the generalizability across different sets of problems (e.g., different types of experiments). Fourth, results were mainly, though not exclusively, based on analyses of verbal protocols. Replicating the major results with different experimental procedures is desirable because the linkages between theoretical constructs and observable variables often are quite speculative in cognitive studies.

A second study was therefore undertaken in order to test and verify the hypotheses generated in the previous study. Power and generalizability were enhanced by recruiting more subjects, manipulating problem familiarity within subjects, using a larger range of problems, and adopting a different experimental task, in addition to verbal protocols.

The purpose of the present study was to determine how the content or quality of the representation of paradigms improves as a function of problem familiarity. Given the importance, demonstrated in previous studies, of accessing the correct paradigm when designing an experiment, the question is how long researchers

are able to choose the correct paradigm when the problem they are confronted with becomes progressively less familiar.

The present study systematically varied problem familiarity for each subject separately, based on the subject's self-reported familiarity with various research domains. Five research domains were chosen for each subject and from each domain a journal article was chosen. One sentence describing the question to be answered in the article was extracted from each of the five articles and presented to the subject. Presumably, the sentence activates a particular paradigm in the subject's long-term memory. The methods of object categorization and feature listing, originally used in categorization experiments by Rosch and associates (e.g., Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976), were used here to determine the content of subjects' paradigms. Choice of the correct paradigm was assessed by asking the subjects to classify a particular research question as being of a certain type. The content of the paradigm retrieved from long-term memory was assessed by asking subjects to write down as many characteristics for this type of research as they could.

The hypotheses are, first, that as research questions become more familiar, subjects will use more specific terms to categorize the questions. For instance, someone unfamiliar with free-recall studies may classify a sentence of this type as being a "psychological study". Someone who is somewhat more familiar may use words such as "cognitive psychological study" or "memory research". An expert may use the words "free recall paradigm". Second, as research questions become more familiar, subjects will list more features. Third, the features listed will become increasingly specific with increasing problem familiarity. With unfamiliar research questions, only general design knowledge can be used and features listed will hence be highly general and applicable to almost any research question. Feature specificity will increase with increasing problem familiarity even when the number of features listed is controlled for. Fourth, with increasing problem familiarity, fewer incorrect features will be listed.

One major confounding variable in this procedure could be that problem familiarity varies together with problem understandability. That is, as problems get less familiar, they also get less understandable, because of the specific terminology used. Although it is arguable whether this is really a confounding variable rather than a variable of interest, it would still be interesting to find out how many and what type of features subjects would list for a novel but understandable research question. Therefore, the fifth research question was identical for each subject and could be considered a control question. The control question did not require any specialized knowledge in order to be understood, since it dealt with the proper layout of calendars.

A comparison between the control question and the question with which each subject was most familiar would show what domain knowledge subjects could bring to bear as a result of their long experience with designing experiments in

their own field of research. We may assume that both questions would be equally understandable. If no difference were to be found between the control question and the most familiar question in the number and type of features listed, then domain knowledge plays a minor role in this task. If subjects would list more and more specific features with the most familiar question than with the control question, then domain knowledge plays a major role in this task. On the other hand, the question with which each subject was least familiar would be less understandable than the control question. A comparison between the control question and the least familiar question would thus show what effect understandability has on problem classification and feature listing. If subjects would be able to list more and more specific features in case of the control question, this would indicate that use of unfamiliar domain-specific terminology limits understandability and presumably hampers accessibility of the appropriate paradigm. If no difference would be found between the control question and the least familiar question, one may conclude that problem understandability does not play an important role in accessing relevant design knowledge.

Research by Rosch et al. (1976) has shown that basic-level concepts such as "chair" contain the most information about the world and are the most differentiated from one another. Categories at the subordinate level, such as "kitchen chair", are not very much differentiated from each other. One way of empirically distinguishing between basic-level categories and subordinate-level categories is by counting the number of new features added at the hypothesized basic level compared with the subordinate level. A feature was defined as new for a particular level if it was not listed at a more inclusive level of abstraction. Rosch et al. (1976) found that the number of new features added at the basic level was significantly more than the number of new features added at the subordinate level. Previous research using the feature listing paradigm with subjects of different levels of expertise has shown that experts in a domain list twice as many features for subordinate level concepts as for basic level concepts (Tanaka & Taylor, 1991). Of all the features listed for subordinate level concepts, approximately half were not listed at the basic level. Hence, a bird expert lists as many novel features for the subordinate concept "sparrow" as for the basic concept "bird". A dog expert, on the other hand, lists fewer novel features for the concept "sparrow" than for the concept "bird", but as many novel features for the concept "beagle" as for the concept "dog". Hence, "extensive knowledge in a domain may result in categories at the level of 'collie' and 'robin' sharing some of the psychological advantages usually attributed solely to categories at the level of 'dog' and 'bird'" (Tanaka & Taylor, 1991, p.478). Similar results have been obtained by Chi, Hutchinson, and Robin (1989). These authors used children who were experts and novices on dinosaurs as subjects. They found that the experts could use their domain knowledge about other dinosaurs (subordinate level concepts) to make inferences about dinosaurs they had never seen before, whereas the novices relied more on knowledge of animals in general (basic level concepts).

Tanaka and Taylor (1991) and Chi, Hutchinson, and Robin (1989) used only two levels of expertise. In the present study, four levels of problem familiarity were used together with a control level. We therefore expected to replicate Tanaka and Taylor's (1991) results for the most extreme levels of problem familiarity. Additionally, we obtained results on the two middle levels of problem familiarity and on the control question. Moreover, it would be interesting to see whether the results obtained by Tanaka and Taylor (1991) on simple stimuli (words such as "dog", "beagle", "animal") could be extended to more complex materials like sentences, or, phrased more generally, whether abstract categories such as paradigms differ from object categories used in the classic research on conceptual hierarchies.

Table I List of predictions for the dependent variables.

Hypothesis	Dependent variable	Prediction
1	Specificity of description of research	High familiarity more specific than low familiarity
2	Total number of features listed	High familiarity more features than low familiarity (control somewhere in between)
3	Specificity of features listed	High familiarity more specific features and fewer general features than low familiarity (control somewhere in between)
4	Number of incorrect features listed	High familiarity fewer incorrect features than low familiarity
5	Number of new features listed	Lowest familiarity < control = highest familiarity
6	Specificity of features listed	More years of overall experience more specific features than fewer years of overall experience

Of particular interest is the position of the control question vis-a-vis the most and the least familiar question. We hypothesize that the control question is comparable to a basic level concept, since it is a research question that can be solved with a minimal amount of domain-specific knowledge by relying only on general design knowledge. The least familiar question would then be comparable to a superordinate level concept and the most familiar question would be comparable to a subordinate level concept. If this hypothesis is correct, we would predict that subjects would list more new features for the control question



compared to the least familiar question, but almost as many new features when compared with the most familiar question.

A final question of interest was whether number of years of overall experience in designing experiments had an effect on problem classification and feature listing. Three experience levels were defined in advance: from three to six years of experience, from seven to ten years of experience, and from eleven years of experience upwards. The hypothesis was that with increasing years of experience, more specific features would be listed.

Table I sums up the hypotheses mentioned above.

## 2 METHOD

### 2.1 Subjects

Thirty-four subjects participated in the experiment. All subjects were working at the TNO Institute for Perception. Experience with designing experiments ranged from three years to thirty years. At the lower end of the experience level were Ph.D. students, while at the higher end experienced researchers in their beginning fifties participated. The subjects had different backgrounds, ranging from psychology to physics to engineering. They were working in the following areas: acoustics, vision, cognitive psychology, performance theory, psychophysiology, thermophysiology, traffic behaviour, training, and motion sickness. All subjects participated voluntarily.

### 2.2 Stimuli

The first step in preparing the stimulus materials was the identification of relevant research areas and obtaining a self-reported familiarity score of each subject on each of the areas. Identification of relevant research areas was accomplished by tracing for each potential subject one or more journals in which they had published or to which they frequently referred to in their reports. In this way, for clusters of three or four subjects one journal was established. If necessary, independent domain experts were consulted on the choice of journals. Most journals were of a theoretical rather than an applied nature. From each journal, one, two or (in one case) three articles were chosen. Care was taken to ensure that the articles were unknown to the subjects, by checking whether they had in recent reports referred to these articles. The Social Sciences Citation Index was consulted to determine the number of times an article was cited in the past two years. All articles were referred to from zero to a maximum of four times (median: 1). This procedure guaranteed that no frequently cited articles were chosen.

In this way, 18 articles were chosen, of which one was the control article. The articles were from the years 1988 ( $N=2$ ), 1989 ( $N=14$ ), or 1990 ( $N=2$ ). For each article, the name of the corresponding research area was determined, again with the aid of independent domain experts of whom most did not serve as subjects in the rest of the experiment (because of a lack of subjects, it was not possible to use independent domain experts for all areas; in only three cases the domain experts also served as subjects). Appendix A lists the articles selected, together with the corresponding research area and the number of citations received.

The resulting names of 17 research areas were presented to the subjects in the form of a questionnaire (the control area was not presented since it would be presented to all subjects later on in the experiment). Subjects were asked to indicate their familiarity with each research area on a scale from 1 to 4. A familiarity score of "1" meant that a subject was unfamiliar with the area and had no idea how to design an experiment in that area; a score of "2" meant that the subject had heard about the area, but could not design an experiment in that area without errors; a score of "3" meant that the subject could design an experiment roughly, since he or she had read about the area once or twice; a score of "4" meant that the subject was highly familiar with the area and could design an experiment quickly and accurately.

In this way, a self-reported familiarity score was obtained from each subject for 17 research areas. On average, 5.3 areas received a familiarity score of "1", 5.7 areas a score of "2", 3.7 areas a score of "3", and 2.3 areas received a familiarity score of "4". In order to assess the reliability of subjects' self-reported familiarity, a subset of 13 subjects was asked to fill in the questionnaire again after a period of five months had elapsed. For these 13 subjects, the mean familiarity score was 2.21 the first time they filled in the questionnaire, and 2.19 the second time, not a statistically significant difference,  $t(220) < 1$ . The Spearman correlation was .78,  $p < .001$ . Cronbach's alpha was .89. Hence, subjects were consistent both in the absolute rating of familiarity as well as in the relative ordering of research areas. The next step was to select the sentences describing each experiment in the articles selected previously. In most cases, this was accomplished by selecting the most critical sentence from the abstract, that is, the sentence that described that "the present article investigated the effects of  $\langle a \rangle$  on  $\langle b \rangle$ ". If this or a similar sentence could not be found in the abstract, then the article itself was read in order to find a sentence of this type.

The next step was to select four sentences for each subject. One sentence came from a research area that had previously received a familiarity score of "1"; the second sentence had received a score of "2"; the third sentence a score of "3" and the last sentence had received a score of "4". One constraint in selecting the four sentences for each subject was that they came from research areas as widely different as possible. For instance, there were two research areas dealing with long-term memory. As far as possible, these two areas were not included for the same subject, unless, of course, these were the only two areas that had received

scores of, for instance, "3" and "4". A second constraint was that the same research area should occur across subjects with at least two degrees of familiarity with this area. For instance, the area of motion vision was included four times for subjects totally unfamiliar with this area, three times for subjects somewhat unfamiliar, four times for subjects moderately familiar, and four times for subjects highly familiar with this particular area. In this way, the same research area was distributed evenly across different levels of familiarity. A third constraint was that the number of research areas included across all subjects should be kept as small as possible in order to keep differences among the sentences selected from the articles as small as possible. In this way, 12 sentences differing in familiarity were selected. One other sentence served as a control sentence and was included for each subject. The resulting 13 sentences are shown in Appendix B.

Booklets with five different sentences were thus constructed for each subject. Sentences were included in random order. On the top of each page a particular sentence was printed. Below, the words "type of research (paradigm):" appeared. Subjects were instructed to describe as specifically as possible in one or two words what type of research the sentence written at the top of the page belonged to. If they knew the particular paradigm used in that sentence, they had to write down the name of the paradigm, else they could use some more general description, such as "psychological research" or "experiment". Below the type of research, the words "characteristics of this research" appeared. Subjects were instructed to write down as many characteristics about this experiment as they could. They were told not to spend more than five minutes on each question. The instructions were given on the front page of the booklet, together with an example that illustrated these instructions. Literal instructions are reported in Appendix C. Each booklet was coded with a number in order to guarantee anonymity to anyone but the experimenter. Of the 38 booklets handed out, 34 were returned.

### 3 RESULTS

For each separate research area, the responses were compiled across subjects. Only if a statement was repeated literally, it was omitted. Complex statements were broken up into separate statements. For instance, if a statement mentioned both a dependent and an independent variable, it was broken up into two statements. If several dependent variables were mentioned in one statement, the statement was left intact. The characteristics mentioned by the subjects were grouped into the following categories: independent variable, dependent variable, control variables, subjects, design, other. The grouping was carried out in order to make comparisons between characteristics within each category easier, and hence increase reliability of scoring. Within each category, characteristics were ordered alphabetically so that statements were not grouped by subject. This was

done in order to avoid context effects in scoring the statements. The characterizations of the type of research were ordered alphabetically too.

For each research area, one or possibly more domain experts were asked to serve as judges and score the subjects' responses. Because of the limited number of domain experts available, most of the judges had participated as subjects. In four of the thirteen research areas, including the control area, the judges had not previously participated as subjects. In order to control for particular biases, an effort was made to have multiple judges score each research area. Seven of the thirteen areas were scored by two judges, the control area was scored by three judges. In this way, 67% of the total of 751 features listed were scored by multiple judges. The judges first scored independently of each other and later discussed their scoring until they reached consensus.

The following scoring system was developed and handed out to the judges. For the type of research, a continuous scale from 0 to 9 was used, with "0" indicating a wrong response and with 1 through 9 indicating a progressively specific response. For the characteristics of the research, a discrete scoring system was developed with four categories:

- "0": wrong characteristic
- "1": superficial reformulation of the sentence, or a correct but highly general characteristic (applicable to every experiment, for instance, "select subjects" or "define ways of measuring variables")
- "2": correct characteristic for this type of experiment but insufficiently worked out and thus still too general
- "3": highly specific and correct characteristic for this type of experiment.

Hence, this scoring system has two underlying dimensions: correctness and specificity. Category "0" indicates a wrong response, whereas the other categories all indicate a correct response. Categories "1" through "3" indicate progressively specific responses. All categories were illustrated with several examples in order to increase coding reliability.

In this way, each statement received a score either from 0 to 9, in case the type of research was scored, or from 0 to 3, in case the characteristics of the research were scored. The statements were then grouped according to whether the sentence the statements belonged to had received a familiarity score of 1, 2, 3, or 4 or whether the sentence was the control sentence.

The following dependent measures were derived from this scoring system:

- 1 specificity of the description of the type of research
- 2 total number of characteristics mentioned
- 3 number of wrong characteristics mentioned (category "0")
- 4 number of characteristics mentioned in categories "1" through "3"
- 5 specificity of characteristics mentioned, with number of characteristics controlled for; for instance, if a subject listed five characteristics, of which three

were classified into category "3", one into category "1" and one into category "0", then a specificity score of  $(3*3 + 1*1)/4 = 2.5$  resulted. Note that the number of wrong responses is excluded from the specificity score.

Univariate and multivariate repeated measures analyses of variance were carried out on these dependent measures. Since the Huynh-Feldt correction indicated that in no case the compound symmetry assumption for univariate repeated measures analyses of variance was violated, the results reported are based on the univariate repeated measures analyses of variance.

We predicted that with increasing years of experience more specific features would be listed. However, no significant effect of level of experience was found on any of the dependent measures. Hence, hypothesis number 6 was rejected. This may have been due to the fact that three to six years of experience already is quite substantial. This between-subjects factor was left out of the remaining analyses, which were therefore within-subjects comparisons only. We will first discuss the results for the four levels of familiarity, and then discuss the effects of understandability by taking into account the control sentence.

### 3.1 Effects of problem familiarity

We expected to find an effect of problem familiarity on the specificity of the words subjects use when asked to classify a sentence as belonging to a particular type of research paradigm (hypothesis 1). The average level of specificity as judged by the domain experts was 2.97, 3.12, 3.09, and 4.71, for familiarity levels 1 to 4, respectively. The overall effect of problem familiarity on level of specificity was significant,  $F(3,99) = 4.74$ ,  $p < .01$ . Planned comparisons showed a significant difference between familiarity levels 1 to 3 versus 4,  $F(1,33) = 10.83$ ,  $p < .01$ . Hence, only for the highest level of familiarity did subjects use specific words to characterize the type of research (99% of the sums of squares of the main effect was accounted for by this planned comparison).

The words subjects used were further investigated by classifying them into two categories:

- 1 very general names such as "experiment", "factorial design", "correlational research" or names of research areas such as "psychological research", "psychophysics", "audiological research", "physiological research";
- 2 names of specific paradigms such as "visual search", "paired-associate paradigm", "signal discrimination experiment, probably 2AFC", "paired comparison of dynamic response", "differential spatial-temporal contrast detection".

For each category, a count was made of the number of times a score from 1 to 8 was assigned to that category (incorrect names receiving a score of "0" were excluded from further analysis). The results showed that 85% of the scores assigned to category 1 ranged from "1" to "3". For category 2, 73% of the scores

assigned to that category were equal to or larger than "4". Given that words with scores from "1" to "3" and words with scores from "4" to "8" form two meaningful categories, I hypothesized that experts highly familiar with a particular type of research would use names of specific paradigms more often than words such as "experiment" or "psychological research". When unfamiliar with particular types of research, the reverse pattern was predicted. The results showed that for familiarity levels 1 to 4, the percentage of subjects using general names was 63%, 58%, 61%, and 35%, respectively. Since two categories were used, the percentage of subjects using specific names accordingly increased from 37%, 42%, and 39% to 65%. The differences between familiarity levels 1 and 3 versus 4 were significant,  $\chi^2(1) = 4.62$  and  $\chi^2(1) = 4.39$ , both  $p$ 's  $< .05$ . The difference between familiarity levels 2 and 4 were marginally significant,  $\chi^2(1) = 3.34$ ,  $p = .07$ . These results confirm the results obtained on the average level of specificity. In addition, they show that two meaningful categories of words can be distinguished.

We predicted that, as research questions become more familiar, subjects will list more features (hypothesis 2). Subjects listed an average number of 4.03, 3.85, 4.47, and 5.15 features for familiarity levels 1 to 4, respectively. The effect of problem familiarity on total number of features listed was significant,  $F(3,99) = 4.36$ ,  $p < .01$ . Hence, with increasing familiarity with a research area, subjects listed more features. Planned comparisons showed a significant difference between familiarity levels 1 to 3 versus 4,  $F(1,33) = 8.69$ ,  $p < .01$ . This result indicates that the total number of features listed only increases with the highest level of familiarity (79.7% of the sums of squares of the main effect was accounted for by this planned comparison).

The total number of features listed can be broken down into correct and incorrect features, and, in case of the correct features, more and less specific features. We predicted that, with increasing problem familiarity, fewer incorrect features would be listed (hypothesis 4). The number of incorrect features listed was .79, .56, .85, and .41, for familiarity levels 1 to 4, respectively. The overall effect of problem familiarity on number of incorrect features was, however, not significant,  $F(3,99) = 1.51$ , ns. Although subjects made somewhat fewer errors on areas they were highly familiar with, this difference did not reach significance. Hence, hypothesis 4 had to be rejected: subjects did not list fewer incorrect features with increasing problem familiarity.

In case of the correct features, the specificity of the features was determined by controlling for the number of features listed. Hence, a higher "specificity score" indicates a higher number of very specific ("category 3") statements, or a lower number of very general ("category 1") statements, or both. We predicted that, with increasing problem familiarity, features listed would become increasingly specific (hypothesis 3). The specificity scores were 1.58, 1.91, 1.88, and 2.27 for familiarity levels 1 to 4, respectively. The overall effect of problem familiarity on level of specificity was significant,  $F(3,99) = 6.94$ ,  $p < .001$ . Hence, with increas-

ing problem familiarity, more domain-specific statements and fewer very general statements were listed. Pairwise comparisons further showed a significant difference between familiarity levels 1 and 2,  $F(1,33) = 5.43$ ,  $p < .05$ , and between levels 3 and 4,  $F(1,33) = 10.02$ ,  $p < .01$ .

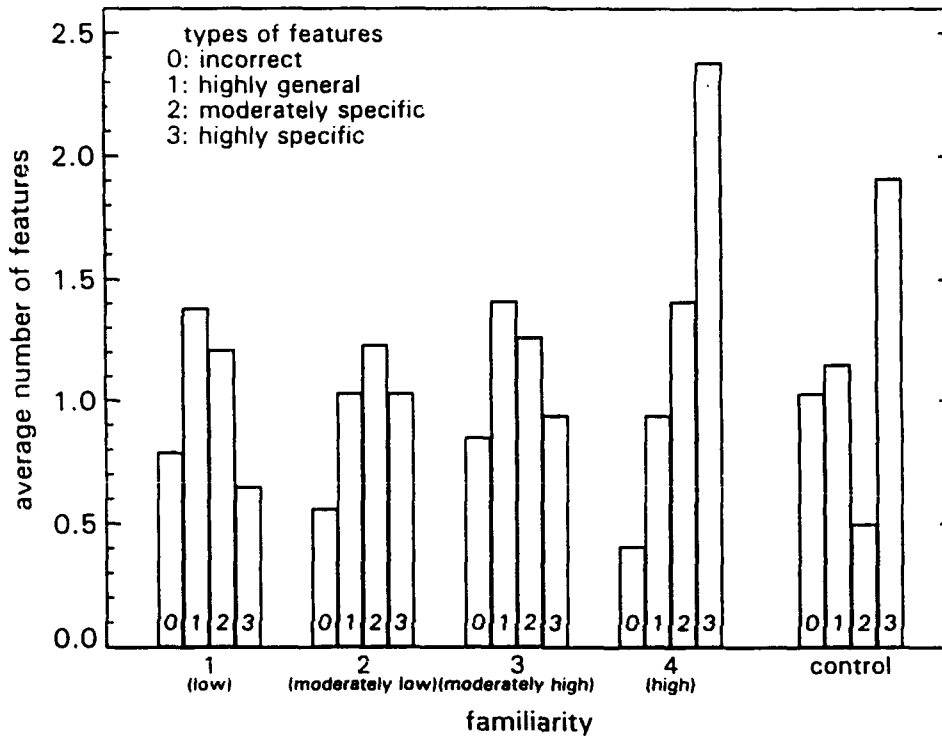


Fig. 1 Average number of four types of features mentioned as a function of familiarity with research areas.

In order to determine more exactly the nature of the increasing specificity of the features listed, the number of statements in categories 1, 2, and 3 were compared. As shown in Fig. 1, the increase in the specificity score can largely be attributed to the large increase in domain-specific ("category 3") statements with problem familiarity 4 as compared to the other levels of problem familiarity. The number of statements in category 3 was .65, 1.03, .94, and 2.38 for familiarity levels 1 to 4, respectively. A planned comparison showed a significant difference between familiarity levels 1 to 3 versus 4,  $F(1,33) = 23.92$ ,  $p < .001$ , and a marginally significant difference between familiarity levels 1 and 2,  $F(1,33) = 3.72$ ,  $p = .06$ . Hence, only when subjects were extremely familiar with a particular research area, were they able to list more highly specific and correct features (95% of the sums of squares of the main effect was accounted for by the planned comparison between familiarity levels 1 to 3 versus 4). Insufficiently worked out and highly general features could be listed in equal numbers for all levels of familiarity.

### 3.2 Effects of understandability

The control sentence was added to the stimuli for all subjects in order to obtain a measure of the type of statements listed if the question is relatively simple and understandable. I expected to replicate Tanaka and Taylor's (1991) finding that experts added almost the same number of new attributes at the subordinate level as at the basic level (hypothesis 5). The subordinate level was hypothesized to be equivalent to the familiarity 4 areas, whereas the basic level was hypothesized to be equivalent to the control area, and the superordinate level to the familiarity 1 area. For our purposes, "number of new attributes" was operationalized as the sum of the number of category 2 and category 3 attributes, for the following reasons.

First, the control sentence was not judged by domain experts, whereas all other sentences were. This was inevitable since the control sentence dealt with a very general research question on which domain expertise is probably non-existent (there are probably no experts on "calendar research"). A possible consequence of this may have been that the scoring criteria were different for the control sentence and the other sentences. In particular, the difference between a category 3 and a category 2 feature may have been less clear in case of the control sentence than in case of the sentences judged by domain experts. Therefore, when comparing the control sentence with the other sentences, it is probably best to combine the number of category 2 and category 3 statements.

Second, incorrect, category 0, attributes were excluded in order to make the results more comparable to the traditional research on conceptual hierarchies, where simple categories (furniture, animals) have been used on which subjects make almost no errors. Third, category 1 attributes were excluded because these are common to all kinds of experiments, and hence are not "new" attributes when listed for a particular kind of experiment. This leaves category 2 and category 3 attributes, which should be combined, since the scoring criteria were probably different for the control sentence and the other sentences for these categories.

The number of new attributes thus defined was 1.85, 2.41, and 3.79 for familiarity level 1, the control sentence, and familiarity level 4, respectively. The control sentence did not significantly differ from familiarity level 1,  $F(1,33) = 2.91$ ,  $p = .10$ , whereas it was significantly different from familiarity level 4,  $F(1,33) = 11.91$ ,  $p < .01$ . These results clearly show that expert knowledge is added primarily at the subordinate level of categorization, since experts added more new features for subordinate-level categories than for basic-level categories. This is a more clear-cut result than reported by Tanaka and Taylor (1991), who found that experts added slightly fewer, and not more, new attributes at the subordinate level than at the basic level. The basic and the superordinate level could not be clearly distinguished, but there was a trend for subjects to list more new features for basic-level categories than for superordinate-level categories.



Familiarity levels 2 and 3 were in between level 1 and the control sentence as far as the number of new attributes listed was concerned (2.26 and 2.21, respectively).

### 3.3 Classification of features

Further analyses of the feature lists were performed to assess whether there were interesting differences in the types of features listed for all familiarity levels and the control sentence. Only category 3 statements were chosen for further analysis, since only these differed substantially across different levels of familiarity. Features listed were classified into the following categories: independent variable, dependent variable, control variables, subjects (number, type), design (within/between subjects), and an "other" category that included features that could not be classified into one of the preceding categories. Table II shows the total number of category 3 statements for the four familiarity levels and the control sentence.

Table II Total number of category 3 statements for the four levels of familiarity and the control sentence

	Fam.1	Fam.2	Fam.3	Fam.4	Control
Independent var.	6	15	14	21	20
Dependent var.	1	10	9	25	27
Control var.	5	3	4	8	0
Subjects	5	2	2	13	4
Design	2	3	2	1	14
Other	3	2	1	13	0

Differences among the proportions of statements in the various categories were tested by a binomial test. Table II clearly shows that the increase in category 3 statements for familiarity level 4 as compared to familiarity levels 1 to 3 is due to statements about the dependent variable ( $p < .01$ ), the control variable ( $p < .05$ ; only for familiarity levels 2 and 3), subjects ( $p < .01$ ), and "other" statements ( $p < .01$ ). Further inspection showed that the "other" statements mainly consisted of hypotheses ( $N = 4$ ) or possible outcomes ( $N = 4$ ) of the experiment. The relatively large number of "design" statements in the control sentence ( $p < .01$  compared with all four familiarity levels) mainly dealt with the issue whether a within or a between-subjects design should be used. The relatively small difference in the number of statements about the independent variable can probably be explained by noting that the independent variable was often explicitly mentioned in the sentence. Interestingly, familiarity level 1 significantly differed from familiarity levels 2 and 3 as far as the number of independent variables ( $p$

$< .05$ ) and the number of dependent variables ( $p < .01$ ) mentioned was concerned. Subjects who were not at all familiar with a research area probably had difficulty comprehending the sentences, and hence extracting the relevant variables from the sentences.

#### 4 DISCUSSION

The main purpose of the present study was to determine how the content of a representation depends on the familiarity with a problem. The content of the representation of a research paradigm was assessed by asking subjects to list features of experiments of varying familiarity. Subjects were presented with a single sentence describing the basic research question that was investigated. They were required to "go beyond" what was explicitly stated in the sentence in order to generate enough specific features. In adding extra information to the sentence, subjects presumably used schemata of varying generality. When they were confronted with a relatively unfamiliar sentence, they had to fall back upon general knowledge about experiments, for instance, that every experiment presents some stimuli to a subject and that a response is measured. If they could not add any more specific knowledge, they would have to rephrase basic elements in the sentence. On the other hand, when subjects were confronted with relatively familiar sentences, they would find it easy to list highly specific features of the particular experiment. In that case, they could presumably use well-structured and elaborated schemata. The purpose of the present experiment was to shed light on the content of those schemata.

The results showed that as subjects were more familiar with a particular research area, they used more specific words to classify the area and they listed more features overall. The extra features listed in case of high familiarity with a research area were almost exclusively highly specific features, dealing with how to measure variables, what number and type of subjects to select, what control variables to use, what hypotheses to test, and what possible outcomes to expect. These features distinguished the experts in a certain area from subjects who were less familiar with that area. A remarkable result was the sharp drop in both the specificity of words used to classify research and the number of specific features listed when moving from familiarity 4 sentences to familiarity 3 sentences. This drop indicates a rather extreme form of domain-specificity of expertise. When confronted with problems only slightly outside their area of expertise, experts must rely upon general design knowledge (category 1) and general knowledge about what are relevant features for the novel area (category 2). Moreover, they use highly general words, such as "experiment" or "psychological research" to classify the novel types of research. In the domain of medicine, Patel, Groen, and Arocha (1990) found that diagnoses were less accurate when an endocrinologist diagnoses a case in cardiology and vice-versa. Kassirer and Gorry (1978) reported that expert nephrologists when diagnosing a case in their

domain of expertise asked fewer questions, mentioned the correct diagnosis earlier, made a firm diagnosis earlier, and maintained a smaller number of active hypotheses than the two physicians who were not expert in the patient's illness. Hence, only the domain experts have the specific expertise required for an accurate diagnosis or for accessing specific design knowledge.

One could argue that instead of having obtained four levels of familiarity, the results show that only two levels of familiarity were sampled, familiarity level 4 versus familiarity levels 1 to 3. However, this conclusion is not valid. On several dependent measures differences were demonstrated between familiarity level 1 versus familiarity levels 2 and 3. Subjects with familiarity level 1 listed significantly fewer specific statements than subjects with familiarity levels 2 and 3. In particular, they listed fewer statements dealing with independent and dependent variables. When we control for the total number of correct statements listed ("specificity score"), subjects with familiarity level 1 still listed fewer specific statements than subjects with familiarity levels 2 and 3. Hence, we can distinguish among three levels of familiarity: low (familiarity level 1), medium (familiarity levels 2 and 3), and high (familiarity level 4).

The reason subjects listed so many insufficiently detailed features with sentences of low familiarity might be that with these sentences subjects may have tried to guess features that they thought were relevant for that particular type of research. All subjects worked at the same institute and regularly hear about each other's work, even if it is remote from their own area. For instance, one may not be familiar with the details of motion vision research, but most subjects who participated in the present experiment at least knew that very few subjects are usually used in motion vision research. Hence, one may classify the category 2 statements as "informed guessing", based on one's general knowledge of what is appropriate in various kinds of research areas. Informed guessing occurred to a far lesser extent with the control sentence, since this was a completely novel area that subjects had never heard about.

The results showed that subjects added more new features at the subordinate level than at the basic level. Although subjects scored higher on the control sentence than on low familiarity sentences, this difference failed to reach significance. This result certainly argues for the importance of domain knowledge, since the sentences on which subjects were experts were presumably equally understandable as the control sentence, yet subjects listed significantly more domain-specific features. Problem understandability does play a small role, however, since there was a trend for subjects to list more domain-specific features for the control sentence than for the sentences that were presumably difficult to understand, although this trend failed to reach significance.

These results replicate and extend the findings of Tanaka and Taylor (1991), and confirm our initial hypothesis that the familiarity level 4 sentences may be viewed as subordinate-level categories. Two independent results may be adduced

to prove this point. The first is the larger number of new features listed for the high familiarity sentences than for all other sentences. The second is the greater use of subordinate-level or highly specific names when identifying highly familiar types of research than when identifying less familiar types of research. Our hypothesis that the control sentence is a basic-level category distinguishable from a superordinate-level category was rejected. In conclusion, it was possible to extend the classic research on conceptual hierarchies of objects in the environment to abstract categories such as paradigms. When experts refer to a particular type of research, they will not use a basic-level name such as "psychological research" or "memory experiment", but rather a subordinate-level name such as "selective attention" or "paired-associate paradigm". Experts thus know a great deal of highly specific information, but only for their own research area. When confronted with novel problems, they have to fall back on general knowledge about experimental design and general knowledge about other types of research. Novel problems are referred to by their basic-level names or even superordinate-level names, such as "experiment", "correlational research", or "factorial design". A novel result of the present study is that the use of basic- or superordinate-level names occurs when subjects are fairly, but not highly, familiar with particular types of research. Previous research by Tanaka and Taylor (1991) only used two extreme levels of familiarity. The present research has used four levels of familiarity and could therefore establish that the domain-knowledge involved in expertise is extremely specific.

Viewed in the context of my previous research (Schraagen, 1990, 1991), the following picture emerges of how experts solve nonroutine or novel problems. When experts are confronted with novel problems, the content of their knowledge is affected such that their performance suffers. This was shown in the previous study (Schraagen, 1990), where design experts performed less well than domain experts but at the same level as beginners, and in the present study where only domain experts highly familiar with a particular type of research area were able to list a large number of features and use names of specific paradigms. This is consistent with the general literature on expert-novice differences that has shown that, in general, experts excel only when they can use their rich domain knowledge.

However, design experts differed from beginners in a previous study (Schraagen, 1990) in the *form* of their problem solving. Their problem solving could be characterized as much more structured than that of the beginners. Experts also used strategies such as mental simulation and progressive deepening, whereas beginners did not. It has been argued that these strategies are available to everyone, novices and experts alike, and that the use of these strategies is a manifestation of the content knowledge that the experts have (Chi & Bjork, 1991). If this is true, then the content knowledge of the design experts has to be different from both the domain experts' and the beginners' knowledge. A possible explanation is that the design experts in the previous study used incorrect paradigms, explaining their low level of performance compared with

that of the domain experts. These paradigms were not used by the beginners, which explains why the design experts were able to use their general strategies of mental simulation and progressive deepening, whereas the beginners were not.

Further evidence for this proposition comes from a recent training experiment (Schraagen, 1991). In this experiment, two groups of novices received a different instruction in how to design experiments. The only difference between the groups was the way their content knowledge was organized: the experimental group received highly structured knowledge in the form of paradigms, whereas the control group received lists of unstructured design principles. The content knowledge itself was identical for both groups, so that no differences in the quality of their solutions were expected. The results confirmed this prediction. Of major interest was the way both groups solved their problem. The experimental group had to switch less often than the control group when designing an experiment. The experimental group selected a paradigm and went on filling in the details, much like experts. The control group had to backtrack more often. Although no evidence was found for more use of the strategies of mental simulation and progressive deepening by the experimental group, these results suggest that availability and use of structured knowledge may lead to more structured problem solving. Strategies such as mental simulation may not automatically be available to everyone, as suggested by Chi and Bjork (1991). It may well be that these are relatively domain-specific strategies that are acquired only after some experience with designing experiments.

A further difference between beginners and experts, apart from domain knowledge and strategy use, lies in the use of general design knowledge and general knowledge about types of research other than one's own specialty area. Beginners have trouble accessing general design knowledge (Schraagen, 1990), and probably lack knowledge about other types of research. The present study has shown that experts frequently resort to these types of knowledge when confronted with problems outside their area of expertise. The present study has also shown that these types of knowledge, and domain knowledge as well, are acquired rather soon after one has specialized in a particular area, given that no differences were found between subjects with three years of experience and subjects with thirty years of experience. A previous study (Schraagen, 1990) already showed that intermediates and design experts used the same strategies and generated designs of comparable quality.

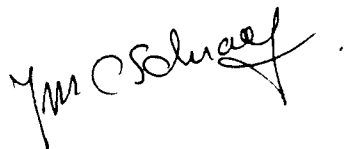
A central tenet of current theories of skill acquisition has been "that high levels of performance reflect specialized domain knowledge that by its very nature is of little or no use in performing tasks in other domains (or even novel tasks within the same domain)" (Holyoak, 1991, p. 307). It is true that domain knowledge is of little or no use when solving novel problems. However, our results have also shown that it is too simple to assume that high levels of performance *only* reflect specialized domain knowledge. Instead, high levels of performance reflect a variety of knowledge and strategies, varying from very general to very specific.

The present study has indicated that this general knowledge is indexed under names such as "experiment" or "physiological research". Presumably, experts possess knowledge about these problem types that they resort to when confronted with novel problems. Since subjects lack more specific knowledge, they have to resort to strategies such as mental simulation in order to solve the problem. However, in the end, when solution quality is assessed, the lack of specific knowledge will always be apparent, no matter how much general knowledge is brought to bear on the problem, and no matter what strategies are used. It seems that when researchers have to design experiments in areas they are unfamiliar with, the experiments they generate will always be of poorer quality than the experiments generated by domain experts.

## REFERENCES

- Chi, M.T.H. & Bjork, R.A. (1991). Modelling expertise. In D. Druckman & R.A. Bjork (Eds), *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press.
- Chi, M.T.H., Hutchinson, J.E. & Robin, A.F. (1989). How inferences about novel domain-related concepts can be constrained by structured knowledge. *Merrill-Palmer Quarterly* 35, 27-62.
- Holyoak, K.J. (1991). Symbolic connectionism: Toward third-generation theories of expertise. In K. Anders Ericsson and J. Smith (Eds), *Toward a general theory of expertise: Prospects and limits*. Cambridge, Mass.: Cambridge University Press.
- Kassirer, J.P. & Gorry, G.A. (1978). Clinical problem solving: A behavioral analysis. *Annals of Internal Medicine* 89, 245-255.
- Patel, V.L., Groen, G.J. & Arocha, J.F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition* 18, 394-406.
- Rosch, E., Mervis, C.B., Gray, W., Johnson, D. & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology* 8, 382-439.
- Schraagen, J.M.C. (1990). How experts solve a novel problem within their domain of expertise. Report IZF 1990 B-14, TNO Institute for Perception, Soesterberg, The Netherlands.
- Schraagen, J.M.C. (1991). The effects of principles and schemata on learning to design experiments. Report IZF 1991 B-16, TNO Institute for Perception, Soesterberg, The Netherlands.
- Tanaka, J.W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23, 457-482.

Soesterberg, July 7, 1992



Drs. J.M.C. Schraagen

# APPENDIX A Articles selected, research areas, number of citations

Ergonomics, 1989 (32), 1373-1389: Alphanumeric and graphic displays for dynamic process monitoring and control

Research area: man-machine interface

Number of citations Jan. '90 - Sept. '91: 1

Journal of Experimental Psychology: Learning, Memory, and Cognition, 1989, 241-245: On the course of forgetting in very long-term memory

Research area: long-term memory (forgetting curve)

Number of citations Jan. '90 - Sept. '91: 4

Journal of Experimental Psychology: Learning, Memory, and Cognition, 1989, 137-146: Bizarre imagery, interference, and distinctiveness

Research area: imagery

Number of citations Jan. '90 - Sept. '91: 3

Journal of Applied Physiology, 1989, 67, 1791-1800: Differential control of forearm and calf vascular resistance during one-leg exercise

Research area: exercise physiology

Number of citations Jan. '90 - Sept. '91: 0

Aviation Space and Environmental Medicine, 1988, 59, 1158-1162: Ocular torsion in upright and tilted positions during hypo- and hypergravity of parabolic flight

Research area: equilibrium and orientation

Number of citations Jan. '90 - Sept. '91: 0

Vision Research, 1989, 29, 1343-1358: Receptive field properties of human motion detector units inferred from spatial frequency masking

Research area: motion vision

Number of citations Jan. '90 - Sept. '91: 0

Journal of the Acoustical Society of America, 1989, 86, 1722-1733: Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation

Research area: auditory signal detection

Number of citations Jan. '90 - Sept. '91: 4



Journal of the Acoustical Society of America, 1989, 86, 1835-1845: The influence of duration on the perception of pitch in single and simultaneous complex tones

Research area: pitch perception

Number of citations Jan. '90 - Sept. '91: 0

Biological Psychology, 1989, 28, 135-148: Reporting of life events, family history of hypertension, and cardiovascular activity at rest and during psychological stress

Research area: medical psychology

Number of citations Jan. '90 - Sept. '91: 0

Biological Psychology, 1990, 31, 107-116: Cortisol reactivity and cognitive performance in a continuous mental task paradigm

Research area: psychophysiology

Number of citations Jan. '90 - Sept. '91: 0

Journal of Experimental Psychology: Human Perception and Performance, 1990, 16, 812-825: Mechanisms of attentional priority

Research area: attention

Number of citations Jan. '90 - Sept. '91: 1

Accident Analysis and Prevention, 1989, 21, 459-468: Relative risk of death from ejection by crash type and crash mode

Research area: traffic safety

Number of citations Jan. '90 - Sept. '91: 1

Ergonomics, 1989 (32), 15-25: Effect of calendar layout on calendar search

Research area: ergonomics of tabular information presentation

Number of citations Jan. '90 - Sept. '91: 0

Journal of Experimental Psychology: Human Perception and Performance, 1988, 14, 638-645: Choosing between-movement sequences: effect of response-choice similarity on the underlying programming operations

Research area: motor behaviour

Number of citations Jan. '90 - Sept. '91: 2

Organizational Behaviour and Human Decision Processes, 1989, 43, 301-335: Misperceptions of feedback in dynamic decision making

Research area: decision making

Number of citations Jan. '90 - Sept. '91: 2

Vision Research, 1989, 29, 1005-1015: Perceived diagonals in grids and lattices

Research area: spatial vision

Number of citations Jan. '90 - Sept. '91: 4

Journal of Applied Physiology, 1989, 67, 438-444: Mechanism of enhanced cold tolerance by an ephedrine-caffeine mixture in humans

Research area: cold physiology

Number of citations Jan. '90 - Sept. '91: 0

Ergonomics, 1989 (32), 27-38: Effects of high visual taskload on the behaviours involved in complex monitoring

Research area: vigilance

Number of citations Jan. '90 - Sept. '91: 1

## APPENDIX B Sentences used as stimuli

- 1 This paper describes two experiments intended to test excitation-pattern models of frequency discrimination by investigating the combined effects of random variations in level and of the addition of a noise designed to mask the upper sides of the excitation patterns of the signals to be discriminated
- 2 The influence of duration on the perception of virtual pitch of complex tones was measured
- 3 The task of this paper was to use a masking technique to isolate families of motion detector units in human vision with the same spatio-temporal properties, and measure their spatial frequency tuning
- 4 The current study examined the relative risk of fatality due to ejection from the vehicle, by crash type and crash mode
- 5 This research evaluated the effectiveness of alphanumeric and graphic display formats for presenting system information in a dynamic process plant environment
- 6 Using unmixed lists, we tested the view that bizarre images would be less susceptible than common (normal) images to interference
- 7 The time course of forgetting in very long-term memory, for events that had occurred from 1 to 15 years ago, was investigated
- 8 Three experiments investigated whether some number of abrupt onsets in a multielement visual display are processed with higher priority than any number of stimuli without abrupt onsets
- 9 In this study we examined whether salivary cortisol, used as an index of stress evoked by the continuous performance of mental tasks, reflected individual differences in cognitive performance
- 10 This study investigated whether relations between stressful life events and cardiovascular activity obtained during periods of rest and stress varied as a function of family history of hypertension
- 11 Four subjects considered resistant to motion sickness were tested in parabolic flight to examine ocular torsion at hypo- and hypergravity
- 12 The purpose of this study was to determine whether blood flow and vascular resistance are controlled differently in the nonactive arm and leg during submaximal rhythmic exercise

- 13 Although the conventional calendar month is formatted as an arrangement of 7 days x 5 weeks, the weeks are sometimes configured as horizontal rows and sometimes as vertical columns, and the day which begins the week is sometimes Sunday and sometimes Monday. The experiment reported here looked at the effects of configuration and beginning day on search performance.

## APPENDIX C Literal instructions for feature listing and sentence classification

On the following pages you will find one sentence taken from a journal article. Some sentences may look familiar, others not at all. The question is to indicate for each sentence to what type of research (paradigm) the research discussed in the article belongs to. Please indicate this as specifically as possible, using jargon. If you find it impossible to indicate this with a particular paradigm, use a more general description of the type of research instead, for instance, "psychological research" or "experiment". Also list as many features of the particular research as you can. You do not need to spend more than 5 minutes on each sentence.

In summary, indicate for each sentence:

- 1 what type of research is discussed here
- 2 list as many features as possible of the particular research.

An example of what is being asked:

The following sentence was taken from the article "Effects of alcohol usage during the first two months of pregnancy on the child's intelligence", from the journal "Social Medicine".

"The goal of the present research was to determine whether the use of alcohol by the mother in the first two months of her pregnancy leads to an increase in mental deficiency compared to pregnancies where the mother does not use alcohol."

Paradigm: longitudinal correlational research

Characteristics of this research:

- operational definition of "mental deficiency": score on a standard IQ-test below 80 at certain age
- matching of mothers on relevant characteristics (age, socio-economic status, area of living: city-rural, IQ parents)
- determination of alcohol usage by self-report mother during pregnancy; verification via partner
- hide questions concerning alcohol usage among other questions
- extensive field research with large (> 1000) number of subjects.

All answers will, of course, be kept highly confidential and without disclosing your name. I will collect this booklet in about a week. Thank you very much in advance for your cooperation.

# REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL) TD 92-0259	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER IZF 1992 B-4
4. PROJECT/TASK/WORK UNIT NO. 787.1	5. CONTRACT NUMBER B92-11	6. REPORT DATE July 7, 1992
7. NUMBER OF PAGES 31	8. NUMBER OF REFERENCES 9	9. TYPE OF REPORT AND DATES COVERED Final
10. TITLE AND SUBTITLE Representation of research paradigms as a function of familiarity with research domains		
11. AUTHOR(S) J.M.C. Schraagen		
12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) TNO Institute for Perception Kampweg 5 3769 DE SOESTERBERG		
13. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) TNO Defence Research Schoemakerstraat 97 2628 VK Delft		
14. SUPPLEMENTARY NOTES		
15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)  The purpose of the present study was to determine how the content or quality of the representation of knowledge of experimental design improves as a function of problem familiarity. The present study systematically varied problem familiarity for each subject separately. Five research domains were chosen for each subject and from each domain a journal article was chosen. One sentence describing the question to be answered in the article was extracted from each of the five articles and presented to the subject. Choice of the correct paradigm was assessed by asking the subjects to classify a particular research question as being of a certain type. The content of the paradigm was assessed by asking subjects to write down as many characteristics for this type of research as they could. Thirty-four subjects participated in the experiment.  The results showed that only when subjects were highly familiar with a research area did they list highly specific features. When confronted with problems only slightly outside their area of expertise, experts must rely upon general design knowledge and general knowledge about what are relevant features for the novel area. The present study has also shown that these types of knowledge are acquired rather soon after one has specialized in a particular area, given that no differences were found between subjects with three years of experience and subjects with thirty years of experience.		
16. DESCRIPTORS  Memory Problem Solving Transfer of Training		IDENTIFIERS  Categorization Expert-Novice Differences
17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
18. DISTRIBUTION/AVAILABILITY STATEMENT  Unlimited availability		17d. SECURITY CLASSIFICATION (OF TITLES)

## VERZENDLIJST

1. Hoofddirecteur van TNO-Defensieonderzoek
2. Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie
- Hoofd Wetenschappelijk Onderzoek KL
3. { Plv. Hoofd Wetenschappelijk Onderzoek KL
- 4, 5. Hoofd Wetenschappelijk Onderzoek KLu
- Hoofd Wetenschappelijk Onderzoek KM
6. { Plv. Hoofd Wetenschappelijk Onderzoek KM
- 7, 8, 9. Hoofd van het Wetensch. en Techn. Doc.- en Inform.  
Centrum voor de Krijgsmacht

Extra exemplaren van dit rapport kunnen worden aangevraagd  
door tussenkomst van de HWOs of de DWO.