Netherlands
organization for
applied scientific
research

TNO-report

TNO Institute for Perception

P.O. Box 23
3769 ZG Soesterberg
Kampweg 5
3769 DE Soesterberg, The Netherlands
Fax +31 3463 5 39 77
Phone +31 3463 5 62 11

IZF 1990 B-14

J.M.C. Schraagen

HOW EXPERTS SOLVE A NOVEL PROBLEM WITHIN THEIR DOMAIN OF EXPERTISE

02

Number of pages: 49

CONTENTS

Report No.:	IZF 1990 B-14
Title:	How experts solve a novel problem within their domain of expertise
Author:	Drs. J.M.C. Schraagen
Institute:	TNO Institute for Perception TNO Division of National Defence Research Group: Cognitive Psychology
Date:	October 1990
HDO Assignment No.:	B89-35
No. in Program of Work:	733.1

SUMMARY

Research on expert-novice differences has mainly focused on how experts solve familiar problems. We know far less about the skills and knowledge used by experts when they are confronted with novel problems within their area of expertise. This report discusses a study in which verbal protocols were taken from subjects of various expertise designing an experiment in an area they were unfamiliar with. The results showed that even when domain knowledge is lacking, experts solve a problem within their area of expertise by dividing the problem into a number of subproblems that are solved in a specified order. The lack of domain knowledge is compensated for by using abstract knowledge structures and domain-specific strategies. The results suggest that experts are confronted with novel problems, they can bring to bear various types of knowledge and strategies that enable them to outperform novices.

Rap.nr. IZF 1990 B-14 Instituut voor Zintuigfysiologie TNO, Soesterberg

Hoe experts een nieuw probleem binnen hun expertisegebied oplossen

J.M.C. Schraagen

SAMENVATTING

Onderzoek op het gebied van verschillen tussen beginners en experts heeft zich tot nu toe vooral gericht op hoe experts bekende problemen oplossen. Veel minder is bekend over de kennis en vaardigheden die experts gebruiken wanneer ze geconfronteerd worden met nieuwe problemen binnen hun vakgebied. In dit rapport wordt verslag gedaan van een studie waarin proefpersonen van verschillende expertiseniveaus hardop denkend een onderzoek moesten opzetten op een voor hen onbekend gebied. De resultaten laten zien dat zelfs wanneer domeinkennis ontbreekt, experts een probleem binnen hun vakgebied oplossen door dat probleem in een aantal subproblemen op te delen en die vervolgens in een vaste volgorde op te lossen. Het gebrek aan domeinkennis wordt gecompenseerd door abstracte kennisstructuren en domeinspecifieke strategieën te gebruiken. De resultaten suggereren dat wanneer experts met nieuwe problemen worden geconfronteerd, ze verschillende soorten kennis en strategieën kunnen gebruiken die hen in staat stellen beter te presteren dan beginners.

1 INTRODUCTION

In the past ten years, research on problem solving has mainly focused on differences in the way experts and novices structure their knowledge (for reviews, see Glaser, 1984; Greeno & Simon, 1988; Van Lehn, 1989). This research has clearly shown that the expert's knowledge base is more abstract, more principled, and more organized for use than the novice's knowledge base.

However, several important questions have been neglected in the research mentioned above. In a recent review on problem solving and reasoning, Greeno and Simon (1988) mentioned as one of the unanswered questions the interactive development and utilization of general and specific skills. For instance, when confronted with novel problems within their domain of expertise, do experts resort to general strategies (or "weak methods") and behave like novices, or do they transfer more task-specific strategies to these novel problems and perform better than novices? A few studies have been carried out on the problem solving skills experts use when confronted with novel problems (e.g., Adelson & Soloway, 1985; Voss, Greene, Post & Penner, 1983; Larkin, 1983). The results of these studies suggest that experts have learned moderately general strategies such as mental simulation that are nevertheless specific to particular domains, for instance software design (Adelson & Soloway, 1985). When they are confronted with novel problems, experts use these strategies to solve these novel problems. Since novices do not use these strategies, they have to search more and hence perform poorer than experts.

Besides using task-specific strategies, a second way in which experts could perform better than novices, when confronted with novel problems, is by using their more abstract and more principled knowledge base. Novel problems could remind experts of previously solved problems that are similar to the current problem in an abstract way. The study by Voss et al. (1983) showed that experts whose domain knowledge was lacking, still came up with more general subproblems than novices. Evidence for the importance of how knowledge is represented also comes from studies of analogical transfer (Gick & Holyoak, 1980, 1983; Holyoak & Koh, 1987; Novick, 1988). In this research area, a distinction is made between structural and surface problem features. Structural features are abstract, whereas surface features are more literal. Novick (1988) has shown that since the representations of experts include both surface and structural features, spontaneous positive transfer occurs in experts' problem solving attempts when the target problem and its analogue share structural features but are superficially dissimilar. Since the representations of novices include only surface features (e.g., Chi, Feltovich & Glaser, 1981; Adelson, 1981), positive transfer does not occur in novices' problem solving when the target problem and its analogue share structural features but are superficially dissimilar. Since research in this area does not typically make use of verbal protocols, it remains unclear what strategies experts use to determine the appropriate structural features in a problem, and what strategies they use to adapt the analogue to the target problem.

In conclusion, although there has been some research on how experts transfer their knowledge to novel problem situations, the interaction between representations and strategies is often left unclear. Mostly, the focus has been on either strategies or representations, but their joint contribution has not been studied in complex, real-world problems. The present study is an attempt to remedy this situation.

The question how experts solve novel problems may be viewed as a question about the transfer between pre-experimental knowledge and performance on a particular task. The question of the transfer of expert knowledge to novel problems is an important one, both for theoretical and practical reasons. Theoretically, questions dealing with the transfer of knowledge and skills have important implications for theories of knowledge representation (Singley & Anderson, 1989). Practically, finding evidence for positive transfer of expert knowledge to novel situations could have educational implications. The strategies and representations used by experts could be made explicit and perhaps successfully taught to novices. A second practical application could be the incorporation of these strategies and representations in expert systems, thereby making them less brittle than they are now.

The rest of this paper is structured as follows. The next section will outline a theoretical framework that will allow us to make general predictions and provide the vocabulary with which to describe the task to be studied here, namely designing an experiment in the area of sensory psychology. This task is described in detail in the section following the theoretical framework. After this task analysis, we are able to derive a model of expert problem solving in this particular task domain. The model is operationalized in terms of a coding scheme for the verbal protocols used for testing the model. In the results section, the model is tested. Finally, the general discussion will consider the implications of the results for the theoretical framework, as well as for some practical issues.

1.1 Theoretical framework

The theoretical framework outlined below contains a number of elements that are derived from a variety of sources (Anderson, 1983, 1987; Laird, Newell & Rosenbloom, 1987; Jansweijer, 1988; Hamel, 1990).

Current theories of cognition (e.g., Anderson, 1983) attach great importance to *hierarchical goal structures*. Goal structures specify what goals have to be accomplished in order to carry out a task. They function as an efficient sequence of steps for carrying out a task. For instance, when solving physics problems, it is efficient to convert all initial data into SI-units (degrees Fahrenheit into degrees Kelvin, etc.). Beginners frequently forget this step, or carry it out in the middle of solving equations, whereas experts have learned to accomplish this goal before solving the equations (Jansweijer, 1988). Hence, goal structures control behavior and provide task decomposition knowledge. That is, they either divide a task into a number of subtasks or they directly solve a subtask by applying domain knowledge. When all the goals are accomplished in the specified order, problem solving follows a structured path. Goal structures are knowledge structures that are initially derived from task instructions and experience with similar problems. With practice on a particular task, goal structures grow more elaborate and more structured.

The goal structure itself is stored in long-term memory (LTM) and is retrieved after the task specifications are understood. The goal structure is deposited in a *limited-capacity working memory* (WM). Only one goal is currently in the focus of attention, but closely linked goals will probably also receive some activation (Anderson, 1983). One of the consequences of this limited capacity is that subgoals cannot be pursued indefinitely when knowledge is lacking, or else the original goal will be forgotten.

The concept of a goal structure cannot by itself explain why experts may have developed *domain-dependent strategies* (or heuristics) that they can use in novel problem situations. Fixed goal structures control behavior in routine problem solving. However, when knowledge is insufficient and an impasse is encountered (Laird, Newell &

Rosenbloom, 1987), a particular goal cannot be accomplished any more. In these situations, experts may have developed heuristics that tell them what to do next. These heuristics dynamically update the goal structure during problem solving, for instance by setting subgoals to repair the impasse. Problem solving is temporarily halted as the requisite domain knowledge is assembled in another problem space.

The final element in the theoretical framework is a structure that contains all the results of problem solving carried out so far. This structure is called the "*problem conception*" (Hamel, 1990), or problem representation. I will assume the problem conception to be schematically organized (cf. Van Lehn, 1989). This schema is a knowledge structure that is selected when the problem description is read, stored in working memory, and gradually elaborated with domain knowledge during problem solving. The problem conception schema is domain-specific yet general at the same time. It is domain-specific because it specifies what domain knowledge should be included in the open slots of the schema. However, it is also general in that the nature of the knowledge and the relations among slots are specified in advance, independent of the particular problem to be solved. By adapting the schema to a problem, missing data are supplied by a process of elaboration, individual data are identified as values of variables, and irrelevant details are ignored. When the schema is successfully adapted, the problem is said to be *understood*. The problem can now be solved using procedural knowledge contained in the schema.

During actual problem solving, the goal structure and the problem conception are intimately connected. The goal structure controls the selection and refinement of the problem conception schema by applying domain knowledge to a task. In turn, the domain knowledge contained in the problem conception schema allows a particular goal to be accomplished. Therefore, an impoverished problem conception schema will lead to less structured problem solving, involving a frequent appeal to subgoals.

The theoretical framework described above allows us to make predictions about what happens when experts are confronted with novel problems.

First, the literature on expert-novice differences has clearly shown that experts have more elaborate goal structures than novices (Jansweijer, 1988). Hence, their problem solving will be more structured than that of novices. On the other hand, experts whose domain knowledge is lacking because they are confronted with novel problems, will show less structured problem solving than experts whose domain knowledge is not lacking.

Second, novices have not had enough experience with a particular task to have developed heuristics as powerful as the experts'. Hence, experts will make use of domain-dependent strategic knowledge when confronted with novel problems, whereas novices have to rely on domain-independent strategic knowledge (or "weak methods"). The domain-dependent strategic knowledge will constrain search to a greater extent than the weak methods. This greater search constraint will prevent the experts from falling back on seemingly random, novice-like problem solving behavior.

Third, experts have a better integrated and more abstract problem conception schema than novices. The use of a well-integrated problem conception schema implies that experts will not suffer from working memory overload as often as novices. When solving novel problems, experts will frequently go over the same goal again and again, because relevant domain knowledge is lacking and has to be assembled in another problem space. Hence, their problem conception schema will be successively refined, showing a pattern of progressive deepening (De Groot, 1978; Kant & Newell, 1984). More important, when confronted with novel problems, experts will be able to use the general elements in their problem conception schema when they adapt the schema to the problem and hence come up with structurally instead of superficially relevant solutions.

The following section will use the concepts defined above in describing the task subjects had to carry out in the present study.

1.2 Designing an experiment in the area of sensory psychology

The problem solving domain investigated in this study is that of designing an experiment in the area of sensory psychology. The following paragraphs will describe the task of designing experiments by both using empirical sources, theoretical analyses, and handbooks.

Designing experiments is an instance of the generic task of design. This classification is based on properties of the input, the expected output, and the nature of the operation taking place to map input to output (Steels, 1990). The input to experimental design is a research question containing specifications. The output is an object, the research plan, that conforms to these specifications. Generic tasks share the same goal structures and the same types of domain knowledge (Chandrasekaran, 1983). Independent research in various domains has found that design tasks are often decomposed into the following subtasks (Brown & Chandrasekaran, 1986; Malhotra, Thomas, Carroll & Miller, 1980; Marcus, Stout & McDermott, 1987; Mittal, Dym & Morjaria, 1986):

1 test specifications for incompleteness or inconsistency
2 generate or extend a partial solution
3 test the adequacy of the solution by matching it with constraints
4 refine the solution by resolving violated constraints

By means of these subtasks, the input is mapped to the output. Some of the pragmatic problems (Steels, 1990) associated with design tasks are the incompleteness of the specifications, the large number of partial solutions possible, and the limited memory available for storing structure. These pragmatic problems determine to a large extent the strategies and types of domain knowledge used by problem solvers. For instance, the incompleteness of the specifications forces the problem solver to test the specifications by validating the data, broadening or restricting the context, classifying the data, or deducing additional features based on class membership. The large number of partial solutions possible implies a structuring of solutions in terms of typical features and not in terms of necessary and sufficient conditions. The limited memory available forces the problem solver to progressively deepen the solution. This general analysis of design tasks will next be applied to design of experiments.

In handbooks on experimental design (e.g., Kerlinger, 1973, p.300), one often finds the following two general goals that together constitute the task of designing an experiment:

1 Answer the research question
2 Control all sources of variance

Based on the task decomposition of the generic task of design, I will assume that the goal of answering the research question is accomplished by understanding the problem, selecting a paradigm, and pursuing that paradigm. Understanding the problem is the equivalent of testing the specifications, selecting a paradigm is equivalent with generating a partial solution, and pursuing a paradigm is equivalent with refining the solution. The goal of controlling all sources of variance is equivalent with testing the adequacy of the solution.

The notion of a paradigm as the knowledge structure that guides experts' problem solving when designing experiments, was derived by analogy with the medical domain. From previous work in the medical domain (e.g., Feltovich & Barrows, 1984), it was clear that medical experts used complex knowledge structures when processing and recalling medical information. Feltovich and Barrows referred to these knowledge structures as "illness scripts". In the authors' words, a clinician "attempts to represent and understand a patient problem by constructing an integrated script or scenario for how the patient's condition came to be, its major points of malfunction, and its subsequent associated consequences" (p.139). In terms of our theoretical framework, the illness script may be viewed as an example of a problem conception schema.

When designing experiments, it is often useful to classify a particular research question as an instance of a more general question that may be solved by some general research plan (Friedland, 1979; Johnson, Nachtsheim & Zualkerman, 1987) or paradigm. For instance, a research question on "how well people are able to remember faces of criminals they have only seen for a short moment" may be classified as an instance of the more general question: "how well are people able to recognize stimuli". This general question then evokes a "recognition paradigm" from memory that specifies what steps have to be taken to answer this question in a scientific way. More specifically, a paradigm is a general research plan containing a specification of the subjects and the independent and dependent variables to be used in the experiment. A paradigm may also contain specifications of the instructions to subjects, the setting where the experiment is carried out, the outcome of the experiment, and control variables (to be discussed in the next paragraph). Usually a subject is first selected, then receives a treatment in the form of an independent variable, and finally a particular aspect (the dependent variable) is measured. Hence, there is a temporal ordering in the elements constituting the paradigm. Since paradigms are applicable in a wide range of situations, they are indexed with respect to fairly general goals. For instance, a recognition paradigm accomplishes the goal of finding out whether someone, when presented with one or more alternatives, is familiar with those alternatives. A multidimensional scaling paradigm may be indexed under: "this paradigm accomplishes the goal of describing a large number of, often perceptual, stimuli into a fewer number of underlying dimensions". Knowledge about paradigms may be considered a catalog of hierarchically organized prototypes, or "skeletal plans" (Friedland & Iwasaki, 1985). The hierarchical way of structuring enables the problem solver to reduce the number of solutions to search for, which is particularly useful in design tasks, as discussed above.

The goal of controlling all sources of variance is accomplished by generating design principles that minimize the error variance and maximize the systematic variance in an experiment. These general goals are accomplished in turn by more specific goals such as experimental control, reliable measurement, using homogeneous groups of subjects, increasing sample size, and using widely different experimental conditions. The goal of experimental control is still fairly general and is achieved by more specific goals such as "avoid carryover effects". This particular goal may be accomplished by counterbalancing conditions. Control of variance is a goal familiar to all students of experimental psychology, and ways of achieving this goal may be found in any textbook on this subject (e.g., Neale & Liebert, 1980). The general design principles may be viewed as constraints against which the partial solution is tested.

One of the aims of our protocol analyses was to identify the different strategies used by subjects whenever they encountered impasses due to a lack of knowledge. In principle, knowledge may be lacking for each of the goals mentioned above. However, I was not interested in problems beginners might have in understanding the problem statement, since in that case they would not even be able to start designing an experiment. I therefore chose a problem that all subjects would in principle be able to understand, viz. a problem that required knowledge of soft drinks and their taste. This choice of problem allowed us to focus on the knowledge and strategies subjects would bring to bear when actually designing an experiment.

The primary interest in this study was in how experts solve novel problems within their domain of expertise. The domain of expertise in this case was designing psychological experiments. In order to identify what is specific for this particular group of experts, the study included subjects with less experience with designing experiments (i.e., beginners and intermediates) and subjects with more domain-specific knowledge (i.e., domain experts). Hence, the other three groups served as controls. For the domain experts, the problem they had to solve was relatively easy, although not trivial. The use of more than two groups of subjects of varying expertise was inspired by the study of Voss et al. (1983). It avoids a problem usually associated with expert-novice studies, namely that experts may be very different from novices in other respects than their greater experience, for instance in intelligence or motivation. By using more groups, the transition from novice to expert could be viewed in a more gradual way, and allowed us to make more comparisons among groups, thereby helping to "unconfound" some of the expert-novice differences.

2 METHOD

2.1 Overview of the methodology used in the present study

The knowledge and strategies used by subjects were assessed by collecting verbal protocols of subjects while designing an experiment. The analysis of verbal protocols requires a coding scheme by means of which statements can be classified into particular categories. In developing a coding scheme, the researcher should follow particular rules (see Ericsson & Simon, 1984). For instance, a coding scheme should not be based on the protocols the researcher is interested in, but rather on a task analysis. Furthermore, the statements used for developing the coding scheme should be scored independently of each other. This study adopted the following procedure:

1) Protocols of subjects solving a similar problem as in this study were segmented into units corresponding to sentences, or, in some cases, larger idea units. Each unit was typed on a card. The resultant deck of 58 cards was given to six other subjects who had not solved the problem but who were familiar with the area of experimental design. Cards were presented to the subjects in a random order, thus ensuring independent scoring of each unit. These subjects were asked to sort the cards into as many categories as they thought appropriate.

2) Categories were reduced by cluster analysis. To this end, similarity matrices were developed based on the categories subjects came up with. Two units received a similarity score of 1 when they were placed in the same category and a score of 0 when they were placed in different categories. These similarity matrices were averaged for all six subjects and analyzed by means of a hierarchical cluster analysis. The results of the cluster analysis showed four categories that were named as follows:

a) understand problem
b) operationalize variables (subjects, (in)dependent variables) c) plan (sequence of events)
d) validity issues (e.g., carry-over effects)

Further analysis showed that these categories could fairly objectively be established by looking for particular key words (e.g., words such as 'identify', recognize', and 'taste' indicated problem understanding; sequences of 'then ... and then' indicated the plan for data collection; words such as 'randomize' and 'counterbalance' clearly indicated validity issues).

Hence, the categories themselves and the attribution of statements to these categories were established by fairly objective procedures, thus ensuring sufficient reliability of coding.

3) Based on a task analysis (see above), these four categories were slightly modified and abstracted. This modification resulted in the following four goals that are *sequentially* accomplished in the task of designing experiments:
   a) understand problem
   b) select paradigm
   c) pursue paradigm
   d) control variance

   This goal structure represents an "expert model" of problem solving in the area of designing experiments.

4) Finally, in order to be able to classify actual protocol statements, a coding scheme was developed. The goal structure mentioned above was extended with the following categories:
   a) evaluation statements, whenever there is insufficient knowledge to choose among two or more knowledge structures;
   b) task-oriented statements, dealing with task requirements, questions to the experimenter, and the evaluation of the task as a whole;
   c) monitoring statements or meta-comments, when subjects report about their own problem-solving processes. These verbalizations are often of limited value, since they do not direct subsequent problem solving behavior (Ericsson & Simon, 1984).

The resulting goal structure for the task of designing experiments is shown in Fig. 1.

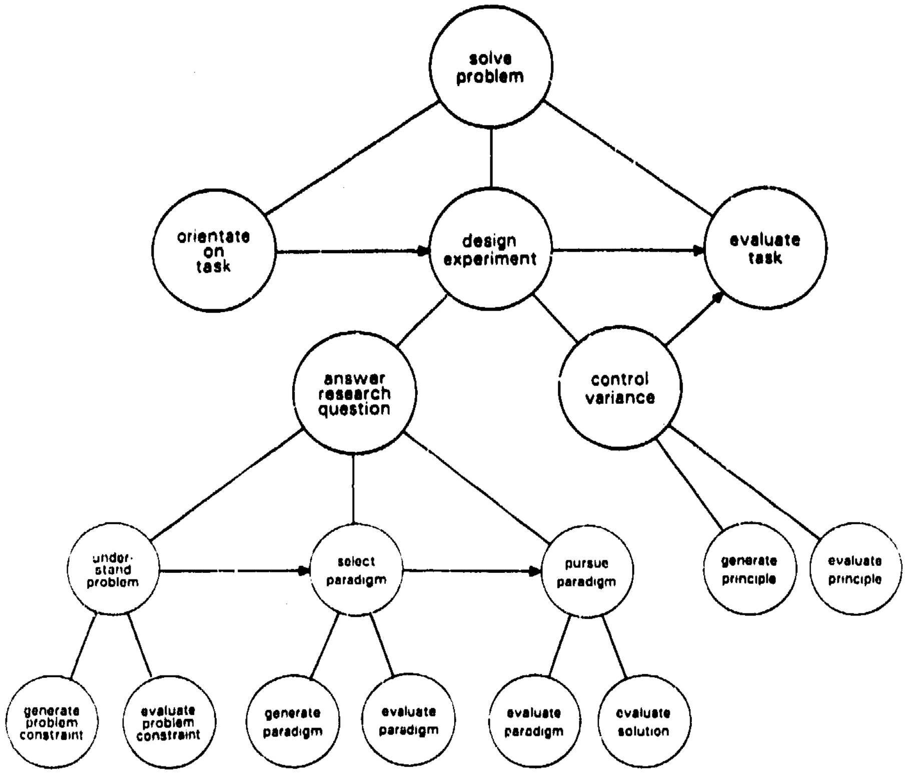

Fig. 1 Goal structure for the task of designing experiments (arrows indicate order in which goals are accomplished).

The coding scheme specifies how the goal structure is manifested in the verbal protocols. Note that the categories in the coding scheme were developed on the basis of a pilot study and not on the basis of the protocols to be discussed in this study. The full coding scheme, with examples from each category, is included in Appendix A. By using the examples and the key words underlined, the experimenter was able to assign statements to categories in a fairly objective way. Hence, no second coder was used to assess inter-rater reliability.

Although, according to the task analysis, the goals are sequentially accomplished, backing up to an immediately preceding goal is allowed, because "activation spreading from the current goal will maintain in working memory the most closely linked goals" (Anderson, 1983, p.161). We may therefore expect to see these associative switches between neighboring goals in verbal protocols.

2.2 Materials

All subjects received the following problem:

> The manufacturer of Coca Cola wants to improve his product. Recently, he has received complaints that Coca Cola does not taste as good any more as it used to. Therefore, he wants to investigate what it is exactly that people taste when they drink Coca Cola. In order to be able to make a comparison with the competitors, Pepsi Cola and a house brand are included in the study as well. The manufacturer has indicated that 'taste' may be defined very broadly in this study. The study will be conducted by a bureau for market research. The manufacturer thinks of the entire Dutch population as the target population.
>
> Please indicate as detailed as possible how, according to you, such a study would look like. You may be able to come up with more than one solution. In that case, do not hesitate and name all of them!

The problem description was deliberately kept vague, in order to bring out differences between subjects in the way they structured the problem, using their knowledge of paradigms. In particular, the problem was vague on the cause of the complaints the cola manufacturer received and on whether the type of study he proposes logically follows from the complaints he has received. The problem description also contained a number of details that subjects may change or abstract from. These details concern the other cola brands, the broad definition of taste, the bureau for market research, and the target population. In reality, researchers are often confronted with questions that are ambiguous, unclear, implicit as far as the main problem is concerned, and loaded with details.

Subjects received the following think aloud instructions on paper (based on Ericsson & Simon, 1984):

> Try to think aloud while performing the task. By this I mean that you tell everything from the moment the task begins until the end of the task. I will ask you to constantly talk aloud during this period. I do not want you to plan ahead what you are going to say. Act as if

> you talk to yourself. It is of the utmost importance that you continue talking. When you fall silent for an extended period of time, the experimenter will ask you to start talking again.

Subjects did not have any trouble thinking aloud while solving the problem.

2.3 Subjects

Four categories of subjects were distinguished:

1) Beginners (Beg): undergraduates majoring in either experimental psychology (N=5) or in methodology (N=4); the beginners' experience with designing experiments was limited to one or two experiments.

2) Intermediates (Int): graduate students in experimental psychology (N=2) or in methodology (N=1); the intermediates' experience with designing experiments was limited to three or four experiments.

3) Design experts (DesExp): subjects with at least ten years of experience in designing experiments in various areas, except in the area of sensory psychology (N=3).

4) Domain experts (DomExp): subjects with at least ten years of experience in designing experiments in the area of sensory psychology (N=4).

2.4 Procedure

Subjects were tested individually in a quiet room at their own or the experimenter's office. The experimenter told them that he was interested in how people of varying levels of expertise designed experiments. Next, subjects were given the problem statement together with the talk aloud instructions. After subjects had read the problem statement, a cassette recorder was started which recorded the subjects' verbalizations. Subjects were allowed to use paper and pencil if they wished to do so. Only two of the design experts made use of these materials. The subjects themselves indicated when they thought they had solved the problem.

2.5 Predictions

Based on our task analysis, the following predictions are made.

First, strategies for pursuing a paradigm will need to accomplish the goal of controlling variance. For novel problems, experienced researchers will use their knowledge of design principles in order to achieve control of variance. Hence, there will be more statements in the "Select design principles" category for the design experts than for the beginners and the domain experts. Second, design experts will switch more often between the categories "Select design principles" and "Pursue paradigm" than beginners and domain experts. Note that both beginners and domain experts also need to accomplish the goal of controlling variance. However, compared with design experts, there will be fewer statements in this category for these two groups. Beginners will have problems retrieving design principles, and domain experts will incorporate these principles directly into their designs, without mentioning them explicitly. Intermediates will perform in between the beginners and the design experts.

Second, overall, domain experts will switch fewer times between categories than design experts and novices, because the domain experts encounter fewer impasses than the two other groups. However, the design experts will conform more to the expert model than the beginners and the intermediates, because of their more abstract problem conception schema and because of their use of domain-dependent strategic knowledge.

Third, paradigms are knowledge structures that are deposited in a working memory with a limited capacity. Paradigms will therefore be successively refined, using a strategy of progressive deepening. Both beginners and domain experts will not use the strategy of progressive deepening. The beginners' knowledge is insufficient for successively adding new information to working memory. The domain experts will, once they have chosen a particular paradigm, pursue that paradigm without having to search for design principles and without having to reread the problem statement. The domain experts will therefore not need to go over the same paradigm again and again. The intermediates will probably have developed rudimentary paradigms, but it is unclear whether they will use progressive deepening.

3 RESULTS AND DISCUSSION

The results section is structured as follows. I will start with some summary statistics on the number of statements in each category of the coding system, the total problem solving time, and the total number of solutions. These results give an overview of some gross differences among the groups. The theoretical framework will provide the categories for discussing the other results. More specifically, the following elements will be discussed: goal structure, strategies for goal attainment and impasse recovery, and problem conception schema.

3.1 Summary statistics

Table I shows the total number of statements in the protocols (with the exclusion of monitoring statements), the total problem solving time for the four groups of subjects, and the total number of solutions (paradigms) mentioned by subjects.

Table I Average total number of statements in protocols, average total problem solving time (in minutes) for the four groups of subjects, and average number of solutions.

	number	time	solutions
Beginners	27	5	1.0
Intermediates	60	9	2.0
Design Experts	66	13	3.0
Domain Experts	68	14	4.2

Clearly, experts came up with more solutions; hence, they took much longer to solve the problem and generated more verbal statements than beginners.

Table II shows the number of statements and the proportion (in brackets) in each category of the coding scheme.

Table II Average number of statements and proportion in each category of the coding scheme for the four groups.

	Beg	Int	DesExp	DomExp
Orientate on task	1 (3%)	1 (2%)	3 (5%)	0 (0%)
Understand problem	5 (17%)	9 (14%)	10 (15%)	19 (28%)
Select paradigm/analogy	3 (10%)	7 (11%)	12 (18%)	17 (25%)
Select design principles	6 (20%)	13 (20%)	15 (22%)	5 (7%)
Pursue paradigm	12 (40%)	30 (47%)	25 (37%)	27 (40%)
Evaluate task	0 (0%)	0 (0%)	1 (1%)	0 (0%)
Monitoring	3 (10%)	4 (6%)	1 (1%)	1 (0.5%)

Since subjects generated more statements with increasing expertise, the analysis on differences between categories was carried out on the proportion of statements within each category. A Kruskal-Wallis Analysis of Variance with level of expertise as grouping variable and the proportion of statements as dependent variable showed a marginally significant difference between the four groups for the category Select design principles ($T=6.40$, $p=0.09$). The remaining categories were not significantly different for the four groups. The first prediction is therefore partly confirmed: the design experts used, across the whole protocol, more design principles than domain experts. Contrary to what was predicted, the beginners made as much use of design principles as the design experts, when the total number of statements generated are controlled for.

3.2 Goal structure

In order to detect an ordering in the goals subjects successively pursued, the switches between the different categories in the protocols were counted. To determine the nature of the switches between the different categories, the three categories: Orientate on task, Evaluate task and monitoring were excluded from further analysis. The reason for the exclusion was that these three categories are not part of the goal structure of interest in this study. Hence, there were four categories left: Understand problem (U), Select paradigm/analogy (SP), Pursue paradigm (PP), and Select design principles (DP).

The switches between the individual statements were classified and counted for each subject. The number of switches was next added for all subjects within one group. The switches *between* categories were tested both against a quasi-random model and against an "expert model", in order to detect whether the data significantly differed from these models. A test against two models gives more confidence in the general pattern of results when, as predicted, one model is accepted and the other rejected. In this case, the random model, but not the expert model, would fit the data of the beginners well, while the reverse pattern is predicted for the expert groups.

The diagonal was excluded from these analyses, because the interest in this study was not primarily in how long subjects would stay in one category. Before presenting the results of the model testing, both models will be discussed in more detail below.

The quasi-random model takes into account the number of items in a particular category and determines the likelihood of going from a particular category to another category. Therefore, the different number of switches between the different groups of subjects is controlled for. If there are more items in a particular category, then chances are higher that a transition will be made to that category, irrespective of the current category.

The expert model is shown in Figure 2.

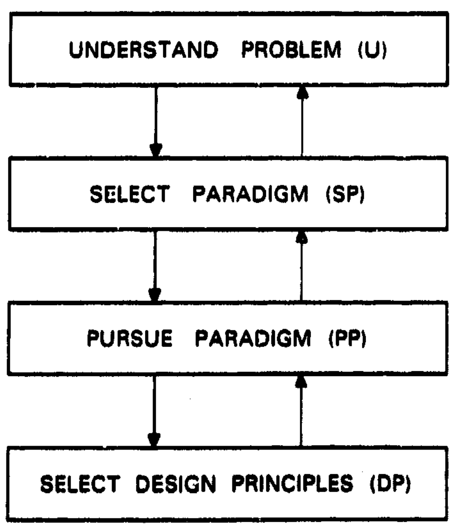

Fig. 2 Expert model.

The expert model only allows switches between immediately preceding and immediately following categories. This yields the following pattern of 'legal' (L) and 'illegal' (I) switches.

Table III 'Legal' and 'illegal' switches according to the 'expert model'.

to from	U	SP	PP	DP
U	-	L	I	I
SP	L	-	L	I
PP	I	L	-	L
DP	I	I	L	-

A constant error parameter was included for every 'illegal' transition. Thus, every illegal switch was considered equally likely. The parameters in the model correspond to weights attached to the categories. The chance of going from one category to the other is proportional with the (relative) weight of the category. There were three parameters in the model that had to be estimated: the error parameter and the parameters corresponding to switches from SP to U and from PP to SP. All other parameters could be derived from these three parameters. Two factors are important when testing the data against the expert model:

1) the 'fit', expressed in a chi-square measure;
2) the magnitude of the error parameter, relative to the other parameters.

Both factors are important, since it is theoretically possible to have a good fit and a high value for the error parameter at the same time. This would be the case when the illegal transitions would all be equal in magnitude and relatively high at the same time. The predictions were that, for the expert groups, first, the data would not significantly deviate from the expert model, and secondly, the error parameter would be low compared with the other parameters. The value of the error parameter was therefore divided by the average value of the other parameters.

The parameters in the models are estimated by minimizing a chi-square function. Hence, the predicted and observed frequencies of switches occurring in the protocols are compared and expressed in a chi-square measure. Table IV shows the results of the parameter estimation.

Table IV Chi-squares (df=5) for the parameter estimates of the four groups.

	Random model		Expert model	
Beginners	6.14	(N.S.)		
Intermediates	20.42	($p<0.001$)	11.72	($p<0.05$)
Design Experts	33.98	($p<0.001$)	9.28	(N.S.)
Domain Experts	21.60	($p<0.001$)	4.56	(N.S.)

The pattern of switches between categories for the beginners did not significantly deviate from the quasi-random model. Both expert groups and the intermediates did significantly deviate from the quasi-random model.

Since the beginners did not significantly differ from the quasi-random model, there was no need to test their data against the expert model. If one would do so, the error parameter would be too high relative to the other parameters. The intermediates significantly deviated from the expert model. The chi-square values for the experts were not significant. The legal parameters were, on average, five times as high as the illegal parameters. The estimated value for the error parameter was 0.15 for the intermediates and 0.13 for the expert groups. These values are very acceptable. Therefore, the conclusion is that the transition data for both groups of experts can be fitted with the 'expert model' described above. The intermediates' data could not be fitted with both models.

In order to test the second prediction, namely that domain experts would switch fewer times between categories than the other groups, the number of switches between categories was divided by the total number of switches. Percentages were calculated since the protocols of the four groups were of unequal length. The percentages for the four groups are shown in Table V.

Table V Percentage of switches between categories for the four groups.

beginners	38%
intermediates	42%
design experts	45%
domain experts	20%

The proportion of switches between categories is much lower for the domain experts than for the other groups. The difference between the four groups is significant (Kruskal-Wallis T=5.91, p=0.05). This confirms our second prediction.

In summary, the results concerning the goal structure yield the following pattern:
- Both groups of experts switched between goals according to the expert model
- The domain experts did not switch as often between goals as the other groups.

The statistics discussed above have given an overall picture of some salient differences between groups in terms of the goal structure. The next section will describe the strategies used by the different groups of subjects to accomplish their goals or to recover from impasses. The focus will be on the design experts, the other groups serving primarily as controls.

3.3 Strategies

Design experts may use deliberate strategies whenever they encounter an impasse and they have to switch to another problem space (category). The following sections will first describe these strategies, illustrating them with protocol fragments where necessary, then describe the criteria used to determine the use of a particular strategy, and finally describe the results in terms of these strategies.

3.3.1 Description of strategies

Strategy 1: Hypothetical reasoning

Hypothetical reasoning is a strategy that is used when the goal is to select a paradigm and there is insufficient knowledge to choose among paradigms. This strategy consists of determining the likely outcome of a particular paradigm and comparing this outcome with what is asked for in the research question. The reasoning process is called 'hypothetical' because the search for a paradigm is carried out in a problem space in which various alternatives are considered as hypotheses and are evaluated before they are actually implemented. Hypothetical reasoning is a form of planning, because the strategy is applied to an abstract search space, in which only the outcomes of paradigms are represented and all other details are ignored.

Design Expert 1 deliberately used this strategy and was aware of its usefulness, as witnessed by the following protocol statements:

> Well, suppose you have done an experiment like that, at least that is always my approach, what do you have? When you have those data, what can you do with them? If you don't know, O.K. I have collected data, but you don't know exactly what to do with those data, then perhaps you should not do the experiment at all.

Strategy 2: Mental simulation

The strategy of mental simulation makes use of the fact that design experts have represented paradigms as scenarios. When subjects tried to fill in the details of a particular paradigm, they would imagine how the experimental procedure would look like. Imagining the procedure often suggested extra information to be included in the paradigm, or difficulties that had to be resolved. The difficulties arose because particular design decisions violated certain validity issues, as specified by certain design principles.

An example where a problem is noted when mentally simulating the procedure, is the following from Design Expert 3's protocol:

> Now you have the problem of: you have three stimuli, you have a subject, you have all controls, and what are you going to do then? ... Well, I think three stimuli are not enough, so you could think about constructing a perceptual space in which you compare those colas with the larger group of soft drinks.

The following example from Design Expert 1's protocol illustrates the use of general design principles:

> And then, secondly, the subject gets a drink, and then I do not know enough about details, whether you have to eat a little bit of bread after that, or wait a minute, or drink something neutral in between, I am not an expert in that area.

In the quote above, the subject interrupts the filling in of the details of the paradigm (after "gets a drink, ..") when he realizes that the subject gets another drink after the first one and that the taste of these two drinks may influence one another. One general design principle is to make sure that the measuring instrument does not change over the course of the experiment (cf. Cook & Campbell, 1979, p.52). In this case, the measuring instrument is the human taster. Design Expert 1 comes up with several ways of preventing this threat to the internal validity of the design, but does not choose among one of them on the ground that he is not an expert when it comes to sensory psychology.

3.3.2 Criteria for identification of strategies

Strategy 1: Hypothetical reasoning

The strategy of hypothetical reasoning occurs before a paradigm is pursued. Subjects tentatively evaluate various paradigms before choosing one. Evidence in the protocols for this strategy comes from the frequent use of words such as "suppose" and "would do".

Strategy 2: Mental simulation

Evidence for the strategy of mental simulation comes from sequences such as 'first...and then...and after that'. For instance: "And then you make a list (...) with a number of dimensions, and then group those dimensions (...) and then let them fill them in (...) and then take those kinds of scores". Just using the words "and then ... and then" is not evidence per se for the use of mental simulation. It might as well be evidence for just summing up the steps in an already stored plan. Mental simulation, on the other hand, means trying out alternatives with the possibility of being corrected. I will restrict the definition of mental simulation therefore to those cases where pursuing a paradigm (indicated by the words "and then ... and then") is interrupted by selecting a design principle.

3.3.3 Description of results in terms of strategies

Strategy 1: Hypothetical reasoning

Design Expert 1 was the only subject who used this strategy. The other design experts immediately chose for a particular paradigm, without extensively evaluating them against other paradigms. The two basic paradigms Design Expert 1 came up with were:

A: pairwise comparisons of colas

B: tasting one cola after the other

There were two versions of both paradigm A and B, and the major task of the subject was to choose between those versions. The two versions are referred to as A1 and A2, and B1 and B2.

An interpretation of the protocol of Design Expert 1, together with impasses and repairs, appears in Appendix B. Appendix B shows that, by using this strategy, Design Expert 1 was able to eliminate two paradigms from his list and ended up by positively evaluating paradigm A2. This paradigm was subsequently pursued. The strategy of hypothetical reasoning was used from the fourth to the eleventh minute in the protocol. This constitutes 50% of the total problem solving time. The remainder of the time was taken up by understanding the problem and pursuing the paradigm. In conclusion, the strategy of hypothetical reasoning enabled Design Expert 1 to constrain his search for possible paradigms.

Strategy 2: Mental simulation

The protocols of all subjects were scored for the use of mental simulation as defined above. The average frequency of use in the four groups of subjects is shown in Table VI.

Table VI Average frequency of use of the strategy of mental simulation for the four groups of subjects.

beginners	0.5
intermediates	3.0
design experts	3.3
domain experts	0.5

The intermediates and the design experts made use of the strategy of mental simulation six times as often as the beginners and the domain experts. A Chi-square test on the total frequency of use of the strategy showed a significant difference between the four groups, Chi-square(3)=12.38, p=0.006. Note that the four groups of subjects are made comparable to each other by carrying out a Chi-square test on the total number of statements in the protocols. The Chi-square test then uses the relative frequencies for the four groups, thus controlling for any differences between the four groups in the total number of statements verbalized.

The results above have already shown that, as predicted, the design experts used more design principles than domain experts. The second part of this prediction stated that design experts would switch more between pursuing a paradigm and use of design principles, and vice versa, than the beginners and the domain experts. The average number of switches between these two categories is shown in Table VII.

Table VII Average number of switches between categories "Pursue paradigm" and "Select design principle" (and vice versa) for the four groups of subjects.

Beginners	3.5
Intermediates	15.7
Design experts	14.7
Domain experts	3.4

The difference between the four groups was highly significant, as indicated by a Chi-square test on the total frequency of switches between the categories, Chi-square(3)=46.48, $p<0.001$. Clearly, the design experts and the intermediates switched more often between the two categories than the other groups. Hence, our prediction is confirmed.

In conclusion, the design experts and the intermediates frequently switched between pursuing a paradigm and selecting and applying a general design principle to that paradigm. Both groups mentally simulated the experimental procedure and frequently interrupted their problem solving whenever a violation of a general design principle was noted. The strategy of mental simulation was less frequently used by the beginners and the domain experts.

The next section will describe the results concerning the final element in our theoretical framework, the problem conception schema.

3.4 Problem conception schema

The third prediction stated that design experts would use the general elements in their problem conception schema when they adapted the schema to the problem, and that they would use a progressive deepening strategy. First, evidence will be shown for the use of general elements in the problem conception schema. Second, evidence for a progressive deepening strategy will be discussed.

The schematizing effect of the problem conception on the ill-structured problem presented to the subjects may be evident from the following elements in the protocols:

- the problem, the research question, or the experiment are categorized, e.g., "a problem on taste", or "consumer research"
- missing information is supplied; this will apply particularly to the important points in the problem description, i.e., the cause of the complaints and the correctness of the manufacturer's research question
- details are abstracted from or changed; the details concern the other cola brands, the broad definition of taste, the bureau for market research, and the target population
- attention is directed to the key elements in the problem formulation, i.e., the phrase "what people taste exactly".

As predicted, the design experts used their problem conception schema for structuring the ill-structured problem they were confronted with. The four elements mentioned above will first be illustrated with relevant protocol segments from the design experts' protocols before turning to a quantitative analysis across the four groups of subjects. All four elements should be clearly identifiable in the protocols as resulting from a particular problem conception schema, rather than being isolated elements that subjects derive from their general world knowledge.

3.4.1 Categorization

Paradigms were chosen quickly on the basis of a structural similarity between the perceived problem and a particular paradigm. Design Expert 2 did not evaluate various paradigms against each other as extensively as Design Expert 1. One of his first statements was:

> I understand that a kind of constraint is that you are thinking of a panel experiment.

This statement indicates that Design Expert 2 had quickly selected a paradigm and saw it as his main task to pursue this paradigm. Design Expert 3 reformulated the problem as follows:

> He wants to know what people taste exactly, so what they take to be the taste of cola. That is of course a pretty vague concept (...) So what you want to measure exactly is: where do my colas fit into a kind of perceptual space of soft drinks.

Design Expert 3 was familiar with Multidimensional Scaling techniques, since he had recently conducted some experiments on the perception of highways using the Personal Construct method and rating scales. It may very well be that reading about perception of taste of colas immediately suggested the same paradigm to him. This suggestion is an example of analogical reasoning in which the target problem and its analogue share structural features (the abstract concept of "perceptual space") but are superficially dissimilar (colas versus highways). Design Expert 2 was also aware of this structural similarity, since he remarked, when trying to come up with a third paradigm:

> Yes, X. [Design Expert 3] has once used a technique, that perception of highways, I would talk to X. how he did that. Because I think the problem is very similar.

Thus, the research question was categorized abstractly by all design experts as falling in the general category of "perceptual experiments", in which an underlying space of dimensions is identified by means of distance ratings between stimuli.

3.4.2 Supplying missing information

During the first minute of his protocol, Design Expert 1 remarked:

> So what they taste exactly, that really is his question. But that does not mean that you have to take that seriously as a researcher, because what does the manufacturer know. Perhaps it is also important to know whether they are able to taste any differences at all. And then pairwise comparisons may be useful.

This fragment shows the use of a particular paradigm, pairwise comparisons, in refining the research question.

3.4.3 Abstraction and changing of details

The following fragment from Design Expert 3's protocol, already discussed above in the context of mental simulation, shows that the number of stimuli is enlarged, because the subject considers using a multi-dimensional scaling paradigm. In this paradigm, a space of underlying dimensions is constructed, using a number of stimuli that is considerably larger than the number of dimensions extracted.

> Now you have the problem of: you have three stimuli, you have a subject, you have all controls, and what are you

> going to do then? ... Well, I think three stimuli are not enough, so you could think about constructing a perceptual space in which you compare those colas with the larger group of soft drinks.

Design Expert 1 also enlarged the number of stimuli when considering a multi-dimensional scaling experiment.

3.4.4 Attention focused on key elements

Design Expert 2 started by saying:

> If the question is really what they taste exactly, then I think you have to use panel research.

This fragment clearly shows that a key element in the problem description triggers a particular paradigm. The same was shown above under the heading "supplying missing information", where Design Expert 1 retrieved a pairwise comparisons experiment after having read the phrase "what people taste exactly".

The beginners invariably immediately translated "taste" into a particular dependent measure, e.g., a rating scale or a questionnaire. There was no evidence for a categorization of the problem or the experiment, i.e., the dependent measure they selected was not part of a larger conceptual structure, but functioned as a goal by itself. Hence, their choice for a dependent measure was based on superficial rather than structural features in the problem statement. One of these superficial features was "the taste of Coca Cola". Reading about the taste of Coca Cola, a lot of beginners were reminded of the "Pepsi challenge", that had been shown on television as a commercial recently. Note that this analogy is actually misleading, since the Pepsi challenge is about preferences for a certain brand of cola, whereas the research question is about "what people really taste", which is a descriptive rather than a hedonic question. Beginners therefore frequently misinterpreted the research question. Interestingly, the beginners could frequently bring to bear a lot of potentially relevant knowledge about soft drinks, e.g., the importance of the image of soft drinks. They failed to incorporate this knowledge into an overall problem conception schema, because they lacked such a schema. Therefore, beginners also did not supply missing information, abstract from details, or focused their attention on the key elements in the problem formulation.

The intermediates did not categorize the problem in the same abstract way as the design experts. Like the beginners, the intermediates started by selecting a particular dependent measure, based upon the fragment "what people taste". The intermediates differed from the beginners, however, in that the dependent measure they chose was part of a paradigm, such as multi-dimensional scaling or self-report questionnaires. Also like the beginners, and unlike the design experts, they did not evaluate their designs against an abstract goal. Two of the intermediates incorrectly specified the goal as: "how well does Coca Cola taste". The intermediates did not check whether the results of their experiments would be of any use to the cola manufacturer. The design experts always checked the use of their results.

The domain experts' choice of paradigm was based on a thorough analysis of the problem statement. The thorough problem analysis is shown by the following quantitative result on the number of statements in the Understand Problem category that occurred before the subjects actually pursued a paradigm.

Table VIII Average number of statements in Understand problem category that occurred before the first statement in the Pursue paradigm category.

Beginners	2.2
Intermediates	6.3
Design Experts	5.7
Domain Experts	10.5

A Kruskal-Wallis Analysis of Variance with level of expertise as grouping variable and the number of statements as dependent variable showed a significant difference between the four groups for the Understand category ($T=9.45$, $p=0.02$). This result indicates that the domain experts devoted more attention to the problem statement before pursuing a particular paradigm than the other groups. The different number of solutions different groups of subjects came up with is not an issue here, since the analysis is carried out on the statements before the first solution is mentioned.

Domain Expert 1 spent a great deal of time analyzing the problem statement. He started his protocol by saying:

> Did the manufacturer translate the problem in the right way? (...) The question is whether those complaints concerning taste do indeed concern the taste. You can have your doubts about that. (...) Look, a complaint, a remark: there are complaints, that asks for: how is the situation, where do those complaints precisely come from. Because I don't think it is right to directly do sensory research.

Domain Expert 1 then went on to enumerate possible causes for the complaints about the taste of Coca Cola: a change in raw materials, natural variations in raw materials, a fault in the process, residues of detergents in cola bottles, fault with the internal quality control, poor marketing and advertisement. Domain Expert 2 mentioned some other possible causes: Coca Cola may have become too expensive, the bottles may have changed in appearance, the "cola-generation" is getting old and switches to other drinks, perhaps due to the introduction of "light beers". These two domain experts generated a large number of hypotheses that might be responsible for the complaints. Most of these hypotheses do not require sensory research, since the problem is not necessarily caused by the taste of cola as such. However, both domain experts went on and assumed that the problem was indeed a sensory problem. From this point on, their problem solving was very similar to that of the other two domain experts, and consisted of retrieving standard sensory paradigms from LTM. The other two domain experts spent much less time analyzing the problem statement, because they assumed that some form of sensory research had to be carried out. This assumption was not made, however, without explicitly questioning the manufacturer's research question. Domain Expert 3 said:

> The first thing I would do is talk to Coca Cola and ask if they really mean what they ask. If they really want to know what people taste exactly, then they can never do that with market research. Then you have to use much more complicated methods, and then I would advise them to set up a descriptive panel.

Besides supplying missing information, domain experts often criticized and changed details in the problem description (e.g., use of the bureau for market research was deemed unnecessary, the target population was defined too broadly, use of the house brand of cola was found illogical). When pursuing a particular paradigm, domain experts often did not refer to cola at all, but rather described general techniques applicable in all kinds of sensory research. This finding indicates that details were abstracted from in the problem conception schema.

In conclusion, only the design experts and the domain experts used the general, structural, elements in their problem conception schema. Use of these general elements allowed them to structure the ill-structured problem they were confronted with. The beginners and the intermediates used superficial features in the problem description to choose an analogy, in the case of the beginners, or a more general paradigm, in the case of the intermediates. Therefore, the first part of the third prediction is confirmed.

3.4.5 Progressive deepening

The second part of the third prediction stated that design experts would use the strategy of progressive deepening, in response to the limited capacity of their working memory. The results of progressive deepening show up in the gradually elaborated problem conception schema.

Progressive deepening is operationalized as follows:

1) the problem solver changes the contents of a particular slot in a paradigm (this change excludes mere repetition of the contents);
2) the slot or its contents have been mentioned before, but not in the immediately preceding protocol statement (this requirement excludes justifying statements).

As defined in the task analysis, the slots in the problem conception schema include the independent and dependent variable, control variables, subjects, and possibly the setting, the outcome, and the instructions to subjects. The following paragraphs illustrate the strategy of progressive deepening in the design experts' protocols.

Progressive deepening was observed in the protocols of all design experts. For example, paradigm B was mentioned four times in all by Design Expert 2, if we ignore the variants for the moment:

1) The first time, B was referred to simply by its generic name "panel research"
2) The second time, the panel research was described more fully by including:
   - the independent variable (three colas)
   - the instructions ("we would like to know what you like or dislike about Coca Cola")

- a control variable (blind, i.e., no brand names visible)
- the dependent variable (description of taste; identification of Coca Cola).

At this stage, the number of subjects and the statistical design are mentioned by the subject, but are left open ("I would not know that right now"). Five elements are mentioned in total the second time the panel research was described.

3) The third time was an extension of what was stated above, and immediately follows it, after the subject has briefly checked the problem statement again. The panel research is now referred to as "free", meaning "with open questions". Design Expert 2 explicitly focuses his attention on this less structured experiment first ("Perhaps it would be good to try to do it in two stages and first allow some open questions"). New categories are added and more items are mentioned with the old ones. The new categories added are:
   - number of subjects (25 to 50)
   - treatment ("have them taste a bit and allow them to go back and forth between colas")
   - statistical analysis ("result is a number of dimensions that are not too clear because of a lot of noise; probably one clear, but uninteresting dimension").

   Items added to old categories:
   - independent variable (three glasses should be coded: a,b,c)
   - controls (balance order)
   - dependent variable (describe differences in taste).

   Eighteen elements are mentioned in total the third time the panel research was described. Thus, comparing the second with the third time the panel research was mentioned, more than three times as many elements are mentioned the third time.
4) The fourth time, paradigm B was referred to as the "more structured" approach. Since this approach uses a different dependent variable than the less structured approach, both approaches cannot be considered extensions of one another.

In this protocol, five slots are successively refined, two of which (the number of subjects and the statistical design) are left open at first, but are explicitly mentioned. The other three slots (independent, dependent, and control variables) are elaborated at several places in the protocol.

The protocols of all subjects were analyzed in this way, and the results are shown in Table IX.

Table IX Average number of successively refined slots for the four groups of subjects.

Beginners	1.5
Intermediates	3.7
Design experts	3.3
Domain experts	1.5

A Chi-square test on the total number of slots showed a significant difference between the four groups for the number of successively refined slots (Chi-square(3)=13.33, p=0.004). The results clearly show that both the intermediates and the design experts successively refined the slots in their problem conception schemata. In contrast, the beginners and the domain experts did not return to slots already filled in. Therefore, the second part of the third prediction is also confirmed.

4 GENERAL DISCUSSION

Generally, the data fit our hypotheses well. First, the main results will be summarized. Next, the results will be interpreted in terms of the theoretical framework developed above. Finally, theoretical and practical implications of the research reported here will be described.

The main results of the present study were:

- design experts' goal structures were much more structured than those of the beginners; the goal structures could not be distinguished from those of the domain experts;
- design experts and intermediates frequently used the strategy of mental simulation; the strategy of hypothetical reasoning was used less frequently and only by the design experts;
- only the design and domain experts' problem conception schemata contained general elements, supplied missing data, helped to focus attention on the important problem features, and abstracted from and changed irrelevant details;
- the problem conception schema of the design experts and the intermediates was gradually elaborated by a strategy of progressive deepening.

What do these results suggest concerning the main question in this study, viz. how do experts solve novel problems within their domain of expertise? The results clearly showed that the design experts in this study did not behave like novices. Instead, their problem solving could be described very well by the same model that described the domain experts' problem solving. Therefore, when knowledge is lacking, the order in which goals are accomplished can remain the same, provided that not too much knowledge is lacking, as was the case with the beginners in this study. When too much knowledge is lacking, the problem solver mainly wanders from one impasse to another, displaying seemingly random search behavior.

The data strongly suggest that the availability of a problem conception schema in the form of a paradigm greatly helps to structure problem solving. When an experimental psychologist is confronted with a novel problem, this problem is first categorized as belonging to a certain abstract category, e.g., "a multi-dimensional scaling problem". This abstract category evokes a paradigm that subsequently guides problem solving, i.e., it helps interpret the problem statement and specifies the general categories (independent variable, etc.) for which information should be obtained. In this sense, experts still exhibit the schema-driven problem solving that characterizes their routine problem solving (e.g., Van Lehn, 1989), even when confronted with novel problems

However, the experimental psychologist may encounter impasses along the way when trying to design an experiment in a novel domain. Several paradigms may be evoked and it may not be clear which one to choose. In this case, the researcher resorts to the strategy of hypothetical reasoning, imagining what the outcome of a paradigm would be and checking this outcome against the problem requirements. When a particular paradigm is chosen, it may not be clear how to fill in the details of that paradigm. In that case, the researcher uses the strategy of mental simulation, imagining how the experiment would look like when it would actually be carried out. When using the strategy of mental simulation, the researcher is frequently reminded of general design principles that apply in this particular case.

The results fit into our theoretical framework as follows. The design experts have developed a task dependent goal structure, that specifies what steps they have to take when an experiment has to be designed. Their general knowledge of paradigms and design principles is indexed with respect to this goal structure, i.e., the relevant knowledge can easily be retrieved whenever a particular goal has to be satisfied. The goal structure remains the same from one problem to another within the same domain. Therefore, this goal structure can be used in solving novel problems. What is lacking in those cases, is the domain knowledge necessary to solve novel problems. This lack of domain knowledge exhibits itself in the protocols as search behavior. Experts constrain their search by using domain-dependent strategic knowledge. Note that "domain-dependent" in this case refers to the domain of experimental design and not the domain of, for instance, sensory psychology. Strategies such as hypothetical reasoning and mental simulation are more generally applicable than in the domain of sensory psychology alone. Therefore, they can be considered as the most general strategies within the domain of designing experiments.

The limited capacity of working memory imposes severe limits on how many goals can be accomplished at once, and how many subgoals can be kept active. When solving novel problems, the domain knowledge often has to be assembled in various problem spaces by a lengthy search process. The design experts in this study often chose not to go into too much detailed search. Instead, they preferred to keep a global picture of the complete paradigm active in working memory. That is, they went over the same paradigm again and again, leaving details open at first, but gradually adding more detail. For instance, they started by referring to a paradigm by its name; when they returned to it, they tried to find a general value for the independent and dependent variable; these general values were subsequently more specified and other elements were considered as well (e.g., control variables, number of subjects, statistical analysis). In short, the design experts used the strategy of progressive deepening.

Neither beginners nor domain experts used these strategies for constraining their search. Beginners only used less successful strategies, which did not result in the retrieval of the knowledge required for solving the problem. Domain experts did not have to use any strategies, since they did not encounter any impasses. Interestingly, the intermediates frequently used the strategy of mental simulation, in conjunction with design principles. They also progressively deepened their paradigm. These results weakly suggest that knowledge of sufficiently detailed and integrated paradigms that the intermediates already possessed enabled them to accomplish their goals in an "expert-like" way. The intermediates' paradigms were not yet as abstract as the design experts', which sometimes resulted in the selection of an incorrect paradigm. In short, the intermediates' *form* of reasoning was similar to that of the experts', but their *content* of reasoning differed.

These results have the following implications for those theories of cognitive skill acquisition (e.g., Anderson, 1987) that place a great deal of emphasis on highly domain-specific knowledge in experts' cognitive skills. First, this study has shown that experts have a flexibility that goes beyond mere domain-specific knowledge. When this knowledge is lacking, experts can still outperform novices by making use of more abstract knowledge and strategies. Current theories of cognitive skill acquisition have mainly focused on the distinction between so-called "weak" and "strong" methods. The present study has shown that there may exist methods of intermediate generality, such as hypothetical reasoning and mental simulation. Exactly how general these methods are is a matter for further research. Second, the present study has indicated how these strategies make use of the abstract knowledge. For instance, abstract knowledge of paradigms contains information about the general outcome of the paradigm. This information can be used by the strategy of hypothetical reasoning. Also, experts have represented the abstract categories in their paradigms in a temporal order. The strategy of mental simulation makes use of this temporal ordering. Third, the results of the present study suggest that a reorganization of declarative knowledge in terms of organized schemata occurs after only limited experience with problem solving. The transition from beginner to intermediate in this study may be viewed as an instance of this reorganization. Since the intermediates were more like the design experts than like the beginners on most relevant measures, this result suggests that the initial reorganization of knowledge plays a much larger role than has been assumed until now.

Practically, these results may have interesting educational implications. Strategic design knowledge, in the form of a goal structure, is not taught explicitly in courses on experimental design, which may be part of the reason why the beginners in this study had so much trouble coming up with a good design. This strategic knowledge is now derived from practice in designing experiments. It may be interesting to try to convey this strategic knowledge for the task of designing experiments to students. Presenting students with high-level goal structures could reduce their working memory load. Studies such as those by Schoenfeld (1979) in the domain of mathematics have indicated that, under appropriate conditions, strategic knowledge can be taught successfully. Besides strategic knowledge, one could try to convey to students the existence of broad classes of research questions and broad classes of answers in terms of paradigms. A training study in which the acquisition of knowledge about paradigms and strategic knowledge would be separately manipulated, could perhaps answer the question why the beginners in this study did not use any of the strategies the design experts used.

A second practical application lies in the area of expert systems. Most expert systems nowadays are competent in very narrow domains. This limited competence makes them very sensitive to slight changes in input data. At the same time, these systems cannot transfer their knowledge to other domains and lack explanatory power. Recently, some attempts have been made to develop systems that are more flexible and are better able to explain their reasoning (Larkin, Reif, Carbonell & Gugliotta, 1985; Clancey, 1988). These systems also incorporate the idea of strategic knowledge, represented separately from the domain knowledge. The goal structure and concomitant strategies that this paper described for the task of designing experiments may also generalize to other tasks involving design. For instance, the strategies of mental simulation, hypothetical reasoning, and progressive deepening have been described in domains such as software design (Adelson & Soloway, 1985; Kant & Newell, 1984), architecture (Coyne, Rosenman, Radford, Balachandran & Gero, 1989; Goel & Pirolli, 1989), and engineering (Goel & Pirolli, 1989). It may well be that these different domains share a number of "design strategies" that could be incorporated in a flexible knowledge-based design system, similar to the diagnostic strategies developed by Clancey (1988).

REFERENCES

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory and Cognition 9, 422-433.

Adelson, B., & Soloway, E. (1985). The role of domain experience in software design. IEEE Transactions on Software Engineering, Vol. SE-11, No. 11, 1351-1360.

Anderson, J.R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J.R. (1987). Skill acquisition: compilation of weak-method problem solutions. Psychological Review 94(2), 192-210.

Brown, D.C. & Chandrasekaran, B. (1986). Knowledge and control for a mechanical design expert system. IEEE Computer 19(7), 92-100.

Chandrasekaran, B. (1983). Towards a taxonomy of problem solving types. AI Magazine 4(1), 9-17.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science 5, 121-152.

Clancey, W.J. (1988). Acquiring, representing, and evaluating a competence model of diagnostic strategy. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds), The nature of expertise (pp. 343-418). Hillsdale, NJ: Lawrence Erlbaum.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: design and analysis issues for field settings. Boston: Houghton Mifflin Company.

Coyne, R.D., Rosenman, M.A., Radford, A.D., Balachandran, M., & Gero, J.S. (1989). Knowledge-based design systems. Reading, MA: Addison-Wesley.

Ericsson, K.A., & Simon, H.A. (1984). Protocol analysis: verbal reports as data. Cambridge, Mass.: MIT Press.

Friedland, P. (1979). Knowledge-based experiment design in molecular genetics. In Proceedings of the Sixth International Joint Conference on Artificial Intelligence, Tokyo.

Friedland, P.E., & Iwasaki, Y. (1985). The concept and implementation of skeletal plans. Journal of Automated Reasoning 1, 161-208.

Gick, M.L., & Holyoak, K.J. (1980). Analogical problem solving. Cognitive Psychology 12, 306-355.

Gick, M.L., & Holyoak, K.J. (1983). Schema induction and analogical transfer. Cognitive Psychology 15, 1-38.

Glaser, R. (1984). Education and thinking: the role of knowledge. American Psychologist 39, 93-104.

Goel, V., & Pirolli, P. (1989). Motivating the notion of generic design within information processing theory: the design problem space. AI Magazine (Spring), 18-36.

Greeno, J.G., & Simon, H.A. (1988). Problem solving and reasoning. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey, & R. Duncan Luce (Eds), Stevens' handbook of experimental psychology (pp. 589-672). New York: John Wiley & Sons.

Groot, A.D. de (1978). Thought and choice in chess (2nd ed.). The Hague: Mouton Publishers.

Hamel, R. (1990). Over het denken van de architect. (in Dutch, with an English summary). Ph.D. Thesis, University of Amsterdam.

Holyoak, K.J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. Memory & Cognition 15, 332-340.

Jansweijer, W.N.H. (1988). PDP: an artificial intelligence approach to problem-solving and learning-by-doing in a semantically rich domain (in Dutch, with an English summary). Ph.D. Thesis, University of Amsterdam. Sneldruk Enschede.

Johnson, P.E., Nachtsheim, C.J., Zualkerman, I.A. (1987). Consultant expertise. Expert Systems 4(3), 180-188.

Kant, E., & Newell, A. (1984). Problem solving techniques for the design of algorithms. Information Processing & Management 20, 97-118.

Kerlinger, F.N. (1973). Foundations of behavioral research (2nd edition). London: Holt, Rinehart and Winston.

Laird, J.E.., Newell, A., & Rosenbloom, P.S. (1987). SOAR: an architecture for general intelligence. Artificial Intelligence 33, 1-64.

Larkin, J.H. (1983). The role of problem representation in physics. In D. Gentner & A.L. Stevens (Eds), Mental models (pp. 75-98). Hillsdale, NJ: Lawrence Erlbaum.

Larkin, J.H., Reif, F., Carbonell, J., Gugliotta, A. (1985). FERMI: a flexible expert reasoner with multi-domain inferencing. Cognitive Science 11, 65-100.

Malhotra, A., Thomas, J.C., Carroll, J.M., & Miller, L.A. (1980). Cognitive processes in design. International Journal of Man-Machine Studies 12, 119-140.

Marcus, S., Stout, J., & McDermott, J. (1988). VT: An expert elevator designer that uses knowledge-based backtracking. AI Magazine 9, 95-112.

Mittal, S., Dym, C.L., & Morjaria, M. (1986). PRIDE: An expert system for the design of paper handling systems. IEEE Computer 19(7), 102-114.

Neale, J.M., & Liebert, R.M. (1980). Science and behavior: an introduction to methods of research (2nd edition). Englewood Cliffs, NJ: Prentice-Hall.

Newell, A. (1989). Putting it all together. In D. Klahr & K. Kotovsky (Eds), Complex information processing: The impact of Herbert A. Simon (pp. 399-440). Hillsdale, NJ: Lawrence Erlbaum.

Novick, L.R. (1988). Analogical transfer, problem similarity, and expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition 14, 510-520.

Schoenfeld, A.H. (1979). Can heuristics be taught? In J. Lochhead and J. Clement (Eds), Cognitive process instruction: research on teaching thinking skills (pp. 315-338). Philadelphia, PA.: The Franklin Institute Press.

Schraagen, J.M.C. (1989). How do experts solve unfamiliar problems: a preliminary study. Soesterberg: Institute for Perception, Report Nr. 1989-31.

Singley, M.K., & Anderson, J.R. (1989). The transfer of cognitive skill. Cambridge, MA: Harvard University Press.

Steels, L. (1990). Components of expertise. AI Magazine 11(2), 28-49.

Van Lehn, K. (1989). Problem solving and cognitive skill acquisition. In M.I. Posner (Ed.), Foundations of cognitive science (pp. 527-579). Cambridge, MA.: MIT Press.

Voss, J.F., Greene, T.R., Post, T.A., & Penner, B.C. (1983). Problem-solving skill in the social sciences. In G.H. Bower (Ed.), The psychology of learning and motivation: advances in research theory (Vol. 17, pp. 165-213). New York: Academic Press.

Soesterberg, October 2, 1990

Drs. J.M.C. Schraagen

Appendix A Coding scheme

Orientate on task (O)

O1: task requirements: "so the task is to say what you *have to* do while thinking aloud"; "and you *want* me to do this under time pressure?"

O2: problem: "it's a well-known problem at any rate: it has received quite some attention in the press"

O3: question to experimenter: "I am allowed to write down things, just for myself?"

Understand problem (U)

U1: generate problem constraint: "manufacturer of Coca Cola *wants* to improve his product"; "so what people taste exactly, that is his *question*"

U2: evaluate problem constraint: "but that does not mean that you have to *take that seriously* as investigator"; "may well be that the Pepsi Cola is preferred at a certain moment"

Select paradigm or analogy (SP)

SP1: generate paradigm or analogy: "then I will go on to a more 'difficult' *experiment* on what people exactly taste"; "I understand that a panel *experiment* is a kind of constraint"; "you could do the well-known Pepsi-challenge"; "you could think of some kind of questionnaire"

SP2: evaluate paradigm or analogy: "maybe we should *abandon* that *plan*"; "and then *pairwise comparisons* may be *useful*"; "if you want to investigate with young children, then questionnaires don't get you very far"

SP3: justify paradigm or analogy: "because if they can't do that, then there is not much use continuing"; "a panel experiment is what comes to mind automatically, because if you were going to interview people on what they taste when they drink Coca Cola, then of course you will have nuisance factors such as image and so on"; "I think I would do it with a card system, just because taste is so difficult to scale"

Select design principles (DP)

DP1: generate principle: "they *have to* be able to switch, I think"; "and finally they also *have to* say which one is cola"; "I think you *need* quite a few subjects"; "and then I don't know enough about details whether you have to eat a little piece of bread after that"

DP2: evaluate principle: "perhaps it would be a *good* idea to try to do it in two stages"; "and then it becomes *less interesting* whether they have to identify cola or not"

DP3: justify principle: "I want an overall judgment, because otherwise I cannot deduce which one you prefer"

DP4: leave details of principle open: "balance order and those kinds of technical details, I don't know whether I have to go that far"; "well, that requires some further consideration"; "that panel, yes how large that would need to be, and the statistical design, I would not know that right now"

Pursue paradigm (PP)

PP1: generate solution: I think I would *take* a glass with a color which just doesn't make you see any differences in color between the drinks"; "and so you *give* a questionnaire and you *let* them score"; "and then, secondly, subject *gets* a drink"

PP2: evaluate solution: "just to make it kind of *fun* for the subjects"; "well, with those data you *would be able* to do something"; "perhaps it is even *better* if you end up there"

PP3: recall solution: "I have already said, non-relevant factors are those image things, matters of order of presentation"

Evaluate task (E)

E1: evaluate task: "is this enough, or do I have to go on, have I forgotten something *important?*"; "what other questions are there?"

Monitoring statements: "I just go on for a moment"; "I am thinking about numerous things"; "I presume I do not have to explain that fully"; "let's see, can I think of anything more about that target population".

In the synthetic protocol, certain words are underlined. When coding the protocol, these words may be used as a guide for classifying statements.

For example, evidence for a subject working on the problem formulation is apparent from words such as 'question' and 'goal' and from literal phrases from the problem formulation.

Use of paradigm is indicated by words such as 'experiment', 'plan', 'method', and 'investigation'.

Verbs such as 'have to', 'can', and 'want' very often indicate use of a design principle, e.g., "You have to provide some open dimensions"; "You have to give an instruction like..."; "The first time I want to measure...".

Verbs in the present tense, such as 'let', 'give', 'get', 'do', and 'have' often give an indication of the current state of the paradigm, e.g., "three of those glasses, let them taste, and then let them name"; "You just have three beakers: a,b,c"; "One gets a drink, and you say...".

Adjectives such as 'important', 'good', and 'interesting' indicate evaluative statements, e.g., "but that is probably an uninteresting dimension"; "that's probably not so bad".

Appendix B Interpretation of protocol of Design Expert 1

1	**Understand problem**
1.1	Read
1.2	Recapitulate
1.3	Write
1.4	Summarize
1.5	Criticise
1.6	Generate alternative research question
1.6.1	Select paradigm A1
1.7	Read
1.8	Recapitulate
1.9	Write
2	**Select paradigm**
2.1	Generate paradigms B1 and B2
2.1.1	impasse: insufficient knowledge to choose between paradigms
2.1.2	repair: Read notes
2.2	Generate paradigms B1 and B2
2.2.1	impasse: insufficient knowledge to choose between paradigms
2.2.2	repair: Generalize givens in problem statement
2.3	Generate paradigm A1
2.3.1	impasse: insufficient knowledge to choose for paradigm A1
2.3.2	repair: evaluate paradigms in evaluation problem space
2.3.2.1	evaluate paradigm A1 by hypothetical reasoning (evaluation negative)
2.3.2.2	evaluate next paradigm on list (B2) by pursuing the paradigm
2.3.2.2.1	impasse: insufficient knowledge to pursue paradigm B2
2.3.2.2.2	repair: evaluate by hypothetical reasoning (evaluation negative)
2.3.2.3	evaluate next paradigm on list (A2) by hypothetical reasoning (evaluation positive)
3	**Pursue paradigm**
3.1	Generate paradigm A2: retrieve from LTM
3.1.1	impasse: insufficient knowledge
3.1.2	repair: select design principle (random presentation of stimuli); stop when too much detail is retrieved
3.2	Generate paradigm B2: retrieve from LTM
3.2.1	impasse: insufficient knowledge to choose among alternatives
3.2.2	repair: avoid too much detail: leave decision open
3.3	Generate paradigm B2: retrieve from LTM

REPORT DOCUMENTATION PAGE

1. DEFENCE REPORT NUMBER (MOD-NL) TD 90-3404	2. RECIPIENT'S ACCESSION NUMBER	3. PERFORMING ORGANIZATION REPORT NUMBER IZF 1990 B-14
4. PROJECT/TASK/WORK UNIT NO. 733.1	5. CONTRACT NUMBER B89-35	6. REPORT DATE October 2, 1990
7. NUMBER OF PAGES 49	8. NUMBER OF REFERENCES 41	9. TYPE OF REPORT AND DATES COVERED Final

10. TITLE AND SUBTITLE

How experts solve a novel problem within their domain of expertise

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 Division of National Defence Research
Koningin Marialaan 21
2595 GA DEN HAAG

14. SUPPLEMENTARY NOTES

15. ABSTRACT (MAXIMUM 200 WORDS, 1044 BYTE)

Research on expert-novice differences has mainly focused on how experts solve familiar problems. We know far less about the skills and knowledge used by experts when they are confronted with novel problems within their area of expertise. This report discusses a study in which verbal protocols were taken from subjects of various expertise designing an experiment in an area they were unfamiliar with. The results showed that even when domain knowledge is lacking, experts solve a problem within their area of expertise by dividing the problem into a number of subproblems that are solved in a specified order. The lack of domain knowledge is compensated for by using abstract knowledge structures and domain-specific strategies. The results suggest that experts are confronted with novel problems, they can bring to bear various types of knowledge and strategies that enable them to outperform novices.

16. DESCRIPTORS

Memory
Problem Solving
Transfer of Training

IDENTIFIERS

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 de Hoofdgroep Defensieonderzoek TNO

2. Directie Wetenschappelijk Onderzoek en Ontwikkeling Defensie

3. (Hoofd Wetenschappelijk Onderzoek KL
(Plv. Hoofd Wetenschappelijk Onderzoek KL

4,5. Hoofd Wetenschappelijk Onderzoek KLu

6. (Hoofd Wetenschappelijk Onderzoek KM
(Plv. Hoofd Wetenschappelijk Onderzoek KM

7, 8, 9. Hoofd van het Wetensch. en Techn. Doc.- en Inform. Centrum voor de Krijgsmacht

LEDEN WAARNEMINGS CONTACT COMMISSIE

10. Maj.Ir. W.C.M. Bouwmans
11. Dr. N. Guns
12. KLTZAR D. Houtman
13. Drs. C.W. Lamberts
14. Ir. P.H. van Overbeek
15. Drs. W. Pelt
16. Maj. dierenarts H.W. Poen
17. Drs. F.H.J.I. Rameckers
18. Kol. drs. G.J.C. Roozendaal
19. LTZSD2OC KV Drs. M.B.A.M. Scheffers
20. Prof.Ir. C. van Schooneveld
21. Ir. M. Vertregt
22. Kol. vliegerarts B. Voorsluijs

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