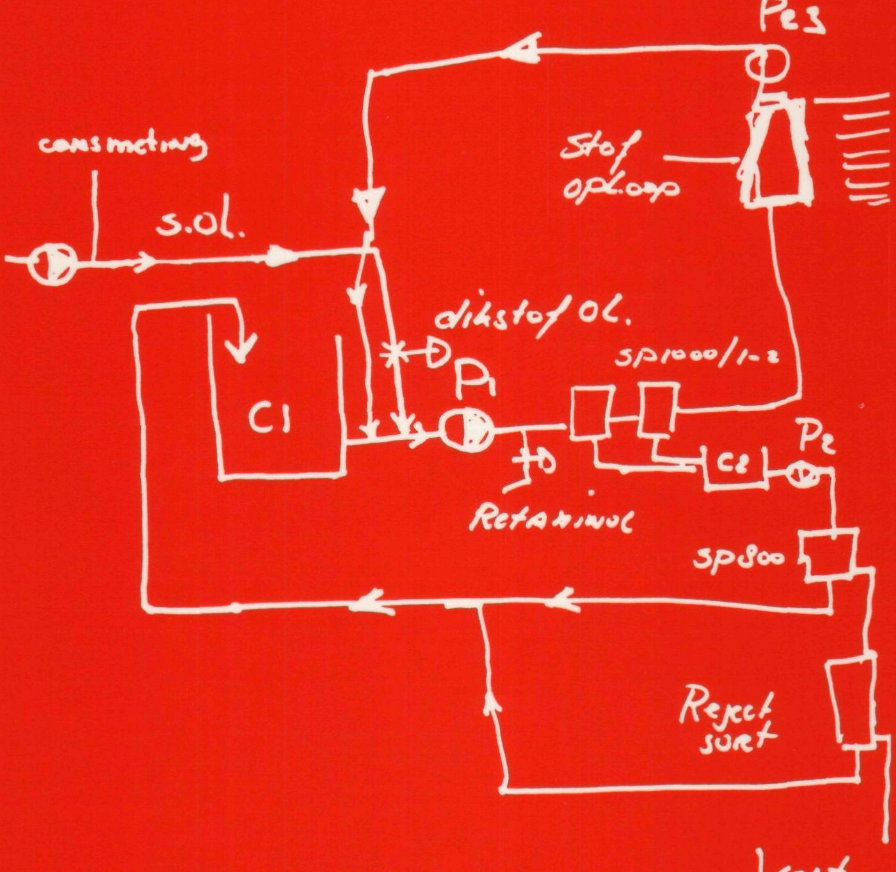
DIAGNOSTIC SKILL IN PROCESS OPERATION

A COMPARISON BETWEEN EXPERTS AND NOVICES



A.M. Schaafstal

RIJKSUNIVERSITEIT GRONINGEN

DIAGNOSTIC SKILL IN PROCESS OPERATION
A COMPARISON BETWEEN EXPERTS AND NOVICES

Proefschrift

ter verkrijging van het doctoraat in de
Psychologische, Pedagogische en
Sociologische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus Dr. S.K. Kuipers
in het openbaar te verdedigen op
donderdag 26 september 1991
des namiddags te 4.00 uur

door

AALTJE MARGRETHE SCHAAFSTAL

geboren op 12 juli 1961
te Eelde

Promotor: Prof. Dr. J.A. Michon

The work presented in this thesis has been performed at the Quality Management Group TNO, Apeldoorn, and the Institute for Perception TNO, Soesterberg.

Cover: part of the drawing of a flowsheet by one of the subjects.

Author's address:

A.M. Schaafstal
Institute for Perception TNO
P.O. Box 23
3769 ZG SOESTERBERG
The Netherlands

Stellingen

behorend bij het proefschrift

Diagnostic skill in process operation
A comparison between experts and novices

van

A.M. Schaafstal

1. Kennissystemen zijn slechts interessant als bijprodukt van kennismanagement: het systematisch behouden, beheren en overdraagbaar maken van kennis.

2. Hoewel domein-specifieke systeemkennis een noodzakelijke voorwaarde is voor het kunnen oplossen van een diagnose probleem, is het zeker geen voldoende voorwaarde voor het toepassen van een goede diagnosestrategie.

   (Dit proefschrift)

3. Gezien de problemen die beginnelingen hebben met het ontwikkelen van een goede taakaanpak, en gegeven het belang van een goede taakaanpak voor een goede taakuitvoering, is het verbazingwekkend om te zien hoe weinig aandacht in het onderwijs wordt besteed aan het aanleren en trainen van strategieën om een taak uit te voeren.

   (Dit proefschrift)

4. Het gegeven dat bij paardrijden zowel de mate van geoefendheid van de ruiter, als de geoefendheid van het paard en de geoefendheid van de combinatie van belang zijn, maakt leren paardrijden tot een lastig trainingsprobleem.

5. Voor cognitiewetenschappers, geïnteresseerd in complexe taakverrichting, is er geen grens tussen zuiver wetenschappelijk en toegepast onderzoek.

6. Juryleden zouden zelf de sport, wier beoefenaars zij beoordelen, moeten beoefenen opdat zij een beter begrip ontwikkelen voor de problemen van technische en emotionele aard waarmee de sporter te kampen heeft.

7. Promovendi die zich in het dankwoord van hun proefschrift excuseren voor hun langdurige "afwezigheid" hadden in plaats daarvan beter voor de aanvang van het onderzoek een cursus projectmanagement kunnen volgen.

8. Voor huishoudens van twee werkende partners die halverwege beider werkplekken gaan wonen geldt niet "gedeelde smart is halve smart", maar eerder "gedeelde smart is dubbele smart".

9. De engelse term "guinea pig" houdt een discriminatie van cavia's in.

10. Het feit dat vaak geen boodschap wordt ingesproken op een antwoordapparaat voor privé-gebruik, geeft aan dat in veel gevallen het relationele aspect van telefonische communicatie belangrijker is dan het informationele aspect.

11. Van vrouwelijke promovendi wordt impliciet verwacht dat zij een stelling over de positie van vrouwen in de samenleving opnemen. Bij deze.

To my mother

Contents

Preface

The research described in this thesis was carried out at the Quality Management Group and the Institute for Perception, both of the Netherlands Organization for Applied Scientific Research (TNO). The cooperation and support of many people during the course of this work have been very valuable to me. I owe them many thanks and I hope there is something I can do for them in return sometime in the future:

My colleagues of the Projectgroep Bedrijfskunde TNO who provided me with sufficient faith in the research to keep going.

The TNO Institute for Perception where the writing was finished. The intellectually stimulating climate made that, all of a sudden, I realized that writing was more fun and easier than I had ever believed before.

The Netherlands Organization for Applied Scientific Research (TNO) which provided extra financial support for this research. To my belief, due to the opportunities to combine research into fundamental issues with applied questions, TNO is an excellent place to carry out interesting research.

My promotor, Prof. Dr. J.A. Michon, for the stimulating discussions we had about many aspects of cognitive science and its application to real-world problems. I will certainly miss those discussions in the future.

The members of my thesis committee, Prof. Dr. G. Mulder, Prof. Drs. J. Moraal, and Prof. Dr. Ir. J.L. Simons for carefully reading the manuscript, and their fruitful comments which came up in discussions afterwards.

Heidi Bogers for scoring the protocols and helping me to refine the coding scheme.

Paul Horstink and Paul Weghorst for playing a major role in developing the simulator used in the experiment reported in chapter Six.

Koos Wolff for producing figures of a professional quality out of rather sketchy drawings.

Special thanks go to Jan Maarten Schraagen and Johan Riemersma who patiently read several versions of the various chapters and every time came up with very good comments, which have greatly improved the manuscript.

I would like to thank Jan Maarten Schraagen with whom I shared many ideas about aspects of the research and writing a thesis. I could not think of a better "sparring partner".

Above all, I am very grateful to the people of Roermond Papier, the paper mill where the research was carried out. Without their participation, there would have been no topic and no research. Thanks to all the subjects, who spent a lot of time on aspects of diagnosis in the experiments. They did so in a very friendly and cooperative way, in times they had to work hard in real troubleshooting situations. Thanks to all other people involved who explained to me more about paper making than I, being not even a novice, could sometimes handle. A special word of thank goes to M. Clumpkens, first of all for the time spent on discussing ideas which I later used in the research, but also for his criticisms on that part of the manuscript which explicitly deals with paper making and its technology. He did his best, the flaws left are mine. Hopefully, the results of this research will eventually become of use to these people.

Summary

Diagnostic skill is one of the most important factors underlying efficient operator behavior. This study deals with the following questions related to diagnostic skill:

- How can diagnostic skill in process operation be characterized?
- Which types of system knowledge play a role in diagnosis, and what is the relationship between system knowledge and problem solving strategies?
- What differences can be found between expert and novice operators with respect to diagnostic skill? This last question is taken as part of the more general question how diagnostic skill develops.

The notion underlying this study, described in chapter One, is that in general, skilled behavior may be attributed to well-structured system knowledge, coupled with efficient problem-solving strategies. Applying this framework to diagnostic skill resulted in the development of a model for a task-level diagnostic strategy as described in chapter Three. The model consists of a description of the different steps in diagnosis, coupled with a specificiation of the ordering of those steps. This model aims at giving an account of actual human behavior with respect to diagnosis, as opposed to an idealized account of problem solving behavior in diagnostic situations that could be used as a specificiation for a knowledge-based system. The model is compared to both two AI-approaches to diagnosis (KADS and diagnosis based on structure and behavior) and the psychological approach taken by, for example, Rasmussen and Rouse (chapter Three). It is concluded that both KADS and approaches to diagnosis based on structure and behavior do not take human intelligent behavior into account as starting point for their model, and there are no explicit considerations with psychological validity. However, they do offer starting points for the development of a model for human strategies applied in diagnosis. The psychological approach taken by Rasmussen and Rouse focusses on only one aspect of diagnosis: strategies for fault finding, which is a subset of the whole process of carrying out a diagnostic task.

The task-level model of diagnosis as described in chapter Three was tested in experiments in a paper mill. The task of the operator in a paper mill, and some general aspects of paper making technology are described in chapter Two. These experiments, using verbal protocols, are reported in chapter Five and Six. It was found that the steps predicted in the model were very much in accordance with the steps that could be deduced from the verbal protocols obtained from expert operators. However, it turned out that the diagnostic strategy applied by novices as well as trainees was rather different from the one employed by expert operators. Both novices and trainees do not systematically make judgments about the seriousness of the situation, and they also "forget" to evaluate their problem solutions. In general, both expert and novice operators do behave according to the predicted ordering of steps. However, the behavior of both expert and novices operators is more in accordance with a depth-first strategy with respect to ruling out possible faults as likely candidates for the problem solution than expected beforehand.

Evidence could be obtained for a, maybe partial, dissociation of the development of a strategy for diagnosis and the availability of system knowledge. First, the experiments reported in chapter Five and Six showed that transfer subjects, expert operators questioned about a rather unfamiliar area of paper making, were able to employ a task-level diagnostic strategy very similar to that of expert operators in that area. Second, the experiments reported in chapter Four and Seven showed that inexperienced operators do not necessarily lack system knowledge. Sometimes they even perform at the same level as their expert colleagues with respect to tasks tapping system knowledge. However, when they have to solve diagnostic problems (chapter Five and Six), this knowledge is not brought to bear, or is insufficient in helping them to solve problems in a systematic way. Thus, although system knowledge may be necessary to support a diagnostic strategy, it certainly is not sufficient to enable the use of a good diagnostic strategy.

Various types of system knowledge could be identified that are used in diagnosis as reported in chapter Five and Six, such as knowledge about the process flow, topographical knowledge, knowledge about paper making, its control, and the process dynamics. The experiments reported in chapter Seven gave evidence for the statement that system knowledge is partially job-dependent: differences were found between the system knowledge of technicians and operators.

Chapter Eight discusses various implications of the research. It is concluded that the task-level model of diagnosis may be useful as an interpretation model in the knowledge acquisition process for the development of knowledge-based systems. The model is assumed to have wider generality than just the domain of paper making. However, only one aspect of diagnosis is discussed, the task-level structure, and to obtain a full model for diagnosis, information about how transitions between the various steps are accomplished should be added, coupled with an investigation about the relation between the diagnostic steps and the underlying system knowledge. The research also provided some insights into how deep (system) and shallow (heuristic) knowledge could be combined in so-called second generation expert systems. One of these insights was that operators are very flexible in the type of knowledge used in a particular situation, and that, presumably dependent on the difficulty of the problem, models at different levels of abstraction are used. Finally, the type of system knowledge used presumably depends on the phase of the diagnostic process: symptom identification may require different types of system knowledge than the identification of possible faults.

Strategy training is currently lacking in many training programs, even though a difference in strategy has been identified in this study as one of the main differences with respect to expert and novice diagnostic skill. With regard to these operator training issues, the task-level model seems an excellent starting point for strategy training.

Samenvatting

Vaardigheid met betrekking tot het kunnen oplossen van diagnose problemen is één van de meest belangrijke factoren die ten grondslag liggen aan een efficiënte taakuitvoering door operators. Deze studie houdt zich bezig met het vinden van een antwoord op de volgende vragen:

- Hoe kan vaardigheid met betrekking tot het kunnen oplossen van diagnose problemen in procesbewakingssituaties worden gekarakteriseerd?
- Welke typen systeemkennis spelen een rol in diagnose, en welke relatie bestaat tussen systeemkennis en probleemoplossingsstrategieën?
- Welke verschillen kunnen worden gevonden tussen beginnende en expert operators op het gebied van diagnose vaardigheden. Deze vraag maakt deel uit van de bredere vraag naar de ontwikkeling van diagnose vaardigheden.

De achterliggende theorie in deze studie, zoals beschreven in hoofdstuk 1, is dat vaardigheden in het algemeen kunnen worden toegeschreven aan goed gestructureerde systeemkennis gekoppeld aan efficiënte probleemoplossingsstrategieën. Dit raamwerk toegepast op vaardigheden met betrekking tot diagnose heeft geresulteerd in de ontwikkeling van een model voor het taakniveau van een diagnose taak, beschreven in hoofdstuk 3. Het model bestaat uit een beschrijving van de verschillende stappen die worden gezet tijdens diagnose, tesamen met een ordening van deze stappen. Doel van het model is het geven van een beschrijving van menselijk gedrag, dit in tegenstelling tot een geïdealiseerde weergave van het probleemoplossingsproces in een diagnose taak die zou kunnen worden gebruikt als specificatie voor een kennissysteem. Het model wordt vergeleken met zowel twee AI-benaderingen met betrekking tot diagnose (KADS en diagnose gebaseerd op structuurbeschrijvingen en gedrag) als een meer psychologische benadering zoals o.a. gehanteerd door Rasmussen en Rouse (hoofdstuk 3). Geconcludeerd wordt dat beide AI-benaderingen niet uitgaan van menselijk intelligent gedrag als vertrekpunt voor de modellering. Bovendien zijn in deze modellen geen expliciete overwegingen gemaakt ten aanzien van psychologische validiteit. Deze modellen vormen echter wel een goed uitgangspunt voor de ontwikkeling van een model van door mensen gehanteerde diagnosestrategieën. De psychologische benadering van Rasmussen en Rouse benadrukt slechts één aspect van diagnose: strategieën voor het localiseren van de fout, hetgeen kan worden opgevat als een onderdeel van het totale proces van diagnostiseren.

Het taakniveau model van diagnose zoals beschreven in hoofdstuk 3 is experimenteel onderzocht in een papierfabriek. De taak van een operator in een papierfabriek en enige algemene aspecten van de technologie van het papiermaken worden daarom beschreven in hoofdstuk 2. De experimenten, waarin gebruikt wordt gemaakt van verbale protocollen, staan beschreven in hoofdstuk 5 en 6. Gevonden werd dat de stappen zoals voorspeld door het model zeer goed overeenkwamen met de stappen die konden worden afgeleid uit de verbale protocollen van expert operators. Het bleek echter dat de diagnosestrategie zoals toegepast door operators in opleiding en beginnende operators sterk afweek van die van de experts. Zowel de leerlingen als de beginners maken geen systematische beoordelingen omtrent de ernst van de situatie, en beide "vergeten" hun oplossingen voor een bepaald probleem te evalueren. In het algemeen kan echter gesteld worden dat zowel beginnende als expert operators binnen de toegelaten stappen van het model blijven. Voor beide groepen geldt overigens wel dat hun gedrag meer in overeenstemming is met een strategie waarbij eerst in de diepte wordt gezocht om mogelijke veroorzakers van een probleem uit te sluiten, dan op voorhand werd verwacht.

Evidentie kon worden verkregen voor een, mogelijk gedeeltelijke, dissociatie tussen de ontwikkeling van een strategie voor diagnose en de beschikbaarheid van systeemkennis. Ten eerste lieten de experimenten, beschreven in hoofdstuk 5 en 6 zien dat transfer proefpersonen, (expert operators ondervraagd over een voor hen relatief onbekend deel van de fabriek), in staat waren om dezelfde strategie toe te passen op de problemen als hun expert collega's. Ten tweede bleek uit de experimenten beschreven in hoofdstuk 4 en 7 dat onervaren operators zich niet anders gedragen vanwege een gemis aan systeemkennis. In sommige gevallen presteren zij zelfs beter dan hun expert collega's op taken die een beroep doen op systeemkennis. Het blijkt echter dat deze kennis niet wordt toegepast in diagnose situaties (hoofdstuk 5 en 6), of onvoldoende steun biedt bij het systematisch oplossen van problemen. Dus alhoewel systeemkennis noodzakelijk geacht wordt ter ondersteuning van een diagnosestrategie, is systeemkennis niet voldoende voor de toepassing van een diagnosestrategie.

Verschillende typen systeemkennis worden gebruikt in een diagnoseproces, zoals beschreven in hoofdstuk 5 en 6, zoals kennis omtrent het stroomdiagram van het proces, kennis omtrent papiermaken en de controle ervan, en kennis over de dynamica van het proces. De experimenten beschreven in hoofdstuk 7 geven aanleiding tot de uitspraak dat systeemkennis voor een deel afhankelijk is van functie: er werden verschillen gevonden tussen de systeemkennis van technici en operators.

Hoofdstuk 8 beschrijft verschillende implicaties van het onderzoek. Geconcludeerd wordt dat het taakniveau model voor diagnose nuttig kan zijn als interpretatie model in het kennisacquisitie proces tijdens de ontwikkeling van kennissystemen. Aangenomen wordt dat het model breder toepasbaar is dan alleen in het domein papiermaken. Hierbij moet worden aangetekend dat slechts één aspect van diagnose is besproken, te weten de structuur van het taakniveau. Om tot een compleet model voor diagnose te komen zou het model enerzijds moeten worden aangevuld met informatie over hoe de overgangen tussen de verschillende stappen plaatsvinden, en anderzijds zou meer inzicht moeten worden verkregen in de relatie tussen de verschillende diagnose stappen en de onderliggende systeemkennis. Het onderzoek heeft ook tot enig inzicht geleid in de vraag hoe diepe kennis (systeemkennis) en oppervlakkige kennis (heuristische kennis) kunnen worden gecombineerd in zgn. tweede generatie expertsystemen. Een van deze inzichten was dat operators zeer flexibel zijn in het type kennis dat wordt gebruikt in een specifieke situatie, en dat, waarschijnlijk afhankelijk van de moeilijkheidsgraad van het probleem, modellen op verschillende abstractieniveau's worden gebruikt. Tenslotte hangt het type systeemkennis waarschijnlijk af van de fase van het diagnose proces: het identificeren van symptomen vraagt om toepassing van andere systeemkennis dan de identificatie van mogelijke fouten.

Training in het gebruik van strategieën ontbreekt momenteel in veel trainingsprogramma's, alhoewel dit onderzoek heeft aangetoond dat een verschil in gehanteerde strategie één van de belangrijkste verschillen is tussen expert en beginnende operators. Het taakniveau model lijkt dan ook een excellent uitgangspunt voor vragen die een rol spelen bij operator training.

Chapter 1

Introduction

1.1 Introduction

Diagnostic skill is one of the most important factors underlying efficient operator behavior. Not surprisingly, it has therefore attracted a lot of attention in recent years, especially in relation to training issues and the occurrence of human error. Much of this research has been carried out on artificial tasks, which may be considered knowledge-lean: they have been constructed in such a way that no "real-world" knowledge can be brought to bear in solving these problems. Hence, the subjects in these studies may be considered novices: these studies are not representative for the range of operator behavior in real-life situations, from which we know that operators are very flexible in their problem solving behavior, that they are very ingenious in using all sorts of available knowledge when applicable, and that diagnostic performance increases with experience. In fact, the acquisition of a cognitive skill in general may well be viewed as the acquisition and use of several types of knowledge (e.g. Anderson, 1987; Gott, 1989). Thus, if we wish to understand what diagnostic skill involves, it is important to study expert behavior in realistic situations, otherwise the experts will be unable to use their knowledge. To appreciate experts' diagnostic skill, it is important to compare their behavior to that of people with less experience in a domain, that is, novice operators. Studying novice operators also offers opportunities for identifying such aspects of diagnostic skill as novices have difficulty with, which could have implications for operator training.

This study deals with the following questions related to diagnostic skill:

- How can diagnostic skill in process operation be characterized?
- Which types of system knowledge play a role in diagnosis, and what is the relationship between system knowledge and problem solving strategies?
- How does diagnostic skill develop or, differently formulated, what differences can be found between expert and novice operators with respect to diagnostic skill?

Strictly speaking, understanding expert-novice differences is just a first step towards understanding how diagnostic skill develops, since it only provides us with a few, extreme points on a developmental scale. However, before we can understand how the learning process evolves, we have to know what limits of the scale can be expected.

This study deals with one domain of process operation, namely process operation in a paper mill, constituting a complex and technical domain. Thus, the study will reflect a number of characteristics of this domain. However, it is assumed that these may be generalized to other domains regarding process operation in technical environments, and it may be that parts of the results are even generalizable to very different diagnostic situations, such as found in the medical domain.

Diagnostic skill is viewed as an example of a cognitive skill and, thus the results should be interpreted in the light of the fundamental question how cognitive skills in semantically rich domains are acquired. In the literature on the acquisition of cognitive skill (e.g. Anderson, 1987; Glaser & Bassok, 1989; Chi, Glaser, & Farr, 1988), it is assumed quite generally that skilled behavior entails a large body of well-structured declarative, or factual, knowledge (knowing that), coupled with efficient problem solving strategies, procedural knowledge (knowing how). However, not much is known about the interaction between declarative and procedural knowledge, and some researchers would even debate whether a distinction between declarative and procedural knowledge is an interesting one (Newell, 1980). Interestingly, a similar interest in the interaction between problem solving strategies (procedural knowledge) and the use of models of devices or installations (declarative knowledge), is also an issue in research on the development of second generation expert systems. By investigating the relationship between system knowledge and diagnostic problem solving strategies, we will be able to contribute to answering these questions.

The results of the study will also have implications of a more applied nature. First, if a concise characterization of diagnostic skill in a real-life situation is made, it will have implications for training issues, since we would than know better what should be trained to obtain good diagnostic skills. Second, a model for diagnostic skill, which will be developed in the course of this study, will be helpful as an interpretation guide for data obtained in the knowledge acquisition process in the development of knowledge-based systems. Finally, on the basis of the results, it may also be possible to hypothesize how operators should be aided in their decision making processes.

The remaining of this chapter is organized as follows. Section 1.2 gives a brief overview of the major findings with respect to expert-novice differences in other domains than paper making. Section 1.3 outlines the approach taken in this study and explains the methodology used, while section 1.4 gives a brief overview of the subsequent chapters.

1.2 Expert-novice differences: a brief overview

During the past several years researchers have been studying the differences between expert and novice problem solvers. Besides the obvious differences in speed and number of errors, researchers have been looking at how experts and novices structure their knowledge and how they approach problem solving.

Some of the earliest work on expert-novice differences was carried out by De Groot (1946) and Chase and Simon (1973a, 1973b) in the domain of chess. They demonstrated that what distinguishes strong (expert) from weak (novice) players are the experts abilities to correctly reproduce large and complex patterns of chess positions after a few seconds of viewing, rather than searching more deeply or broadly than weaker players do. Chase and Simon (1973b) showed that the master chess players' performance was not attributable to a better memory in all situations, since the better recall abilities were restricted to meaningful chess positions: with random patterns of board positions, both experts and novices performed equal. Chase and Simon (1973b) used the concept of *chunk* as a defining unit of knowledge structure, a chunk being a meaningful piece of information that may be identified and used at once. Experts are assumed to have bigger chunks than novices, and the number of chunks will increase with growing experience as well. The ability of experts to perceive large meaningful patterns at a glance within their own domain has been replicated in several other domains, such as the game of Go (Reitman, 1976), reading circuit diagrams (Egan & Schwartz, 1979), reading architectural plans (Akin, 1980), and also in interpreting X-ray plates (Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988). However, this ability should not be attributed to a superior perceptual ability; rather, it reflects an organization of the knowledge base. Programmers, for example, can recall key programming language words in meaningful clusters (McKeithen, Reitman, Rueter, & Hirtle, 1981), and expert programmers can also instantly recognize and recall familiar subroutines (Soloway, Adelson, & Ehrlich, 1988).

Using physics problems, Chi, Feltovich, and Glaser (1981) found that experts used principles of mechanics to organize problem categories, whereas novices built their categories around surface characteristics of the problem, such as elements explicitly mentioned in the problem description. Similar results have been obtained by Weiser and Shertz (1983) in the domain of computer programming. When expert and novice programmers were asked to sort programming problems, the experts sorted them according to solution algorithms, whereas the novices sorted them according to areas of application. These results reflect the more general statement that *experts see and represent a problem in their domain at a deeper (more principled) level than novices; novices tend to represent a problem at a more superficial level* (Glaser & Chi, 1988).

Another important difference between experts and novices is their approach to solving problems. Larkin, McDermott, Simon, and Simon (1980) found that novices solve physics problems using a bottom-up approach: they start with the equations containing the variable and work backward. *Experts start with given information and work forward* because their initial backward solutions become automated with learning. This implies that experts will change to a bottom-up approach only with more complex or new problems, which has indeed been found by Patel, Groen, and Arocha (1990). The second important finding with respect to differences in approach to problem solving is the finding that *experts spend a great deal of time analyzing a problem qualitatively*. Experts typically try to "understand" a problem in the beginning of the problem solving episode, whereas novices plunge immediately into attempting to apply equations and to solve for the unknown (Paige & Simon, 1966). A third common finding, related to the previous one, is that *experts use stronger, more powerful, problem solving strategies* (Anderson, 1983). Finally, *experts have strong self-monitoring skills*. They are much more aware than novices of errors they make, why they fail to comprehend, and when they need to check their solutions (Simon & Simon, 1978; Larkin, 1983).

These results indicate that, when approaching problems, both experts and novices can identify the key information of a problem, but experts have a better understanding of the goals and actions involved in solving a problem and are better at making inferences. Novices have a limited ability to generate inferences and relations not explicitly stated. Thus, problem solving performance at least partly relies on a good problem representation, which will interact with problem solving strategies. Both representations and strategies will develop over time, and become more and more adapted to the domain in which one becomes an expert.

If the findings of these studies on expert-novice differences are translated to the domain of the acquisition of diagnostic skill in paper making, what can be expected? Obviously, like in all the other domains, the experts are expected to be faster and less error-prone than the novice operators. Thus, experts and novices will be *quantitatively* different from each other. This is neither very surprising nor very interesting either. The question is whether any differences can be found that are *qualitative* in nature, similar to the differences described in other domains: a different organization of the knowledge base, different approaches to problem solving, or different problem solving strategies. In short, the question is not whether experts are better than novices, but *why* they are better than novices. Knowledge about the qualitative differences between experts and novices is required to *explain* differences that are quantitative in nature. For example, application of qualitatively different strategies may lead to very different error rates or to differences in proposed solutions. Thus, qualitative differences may provide us with an understanding of quantitative differences. In this study, special attention will therefore be given to *qualitative differences* between experts and novices.

1.3 Approach of the study

This study deals with diagnostic skill in process operation, and in particular, with the development of diagnostic skill in paper making. As pointed out before, the study heavily relies on characterizing expert-novice differences. Since expertise is limited to a very narrow domain, it is important to choose situations for observing expert behavior that closely match realistic and usual situations in the domain. Thus, it was decided that the experiments reported in this study would be carried out in a paper mill. Apart from the fact that this is a very convenient way of matching experimental situations to "normal" situations, it has the additional advantages that it makes it easier for operators to participate and that there are more operators available. In addition to this, it was decided to carry out experiments in the control room itself, when needed or convenient. The disadvantages of experimenting in such a situation are plenty: it may be noisy, experiments get easily disrupted, and it takes a long time for the experimenter to get sufficiently acquainted with the domain. However, the advantages are certainly outweighing the disadvantages. First, experimenting in the paper mill itself turns out to be the only reasonable and practical way to let operators participate in experiments. Second, participating in experiments should not be perceived as a real hindrance for subjects. Since subjects are hard to get by, every care should be taken to ensure that the experiments are as pleasant as possible. This implies that sometimes experiments have to take place in the control room, especially when expert subjects are run, at moments that the manufacturing process is in a stable situation. In those situations, the operator is only involved in some regular monitoring, and there is spare time for other activities, for example for participating in an experiment. If one manages to run experiments during these periods, the operator will still be responsible for the process and the crew members, but is also able to participate, thereby avoiding problems of replacement during his absence. In actual fact, this has certainly enhanced the willingness of operators to participate and the willingness of the managers to allow experiments to be conducted in their mill. However, there still is a rather high likelihood that the experiment has to be stopped due to an unforeseen problem that the subject has to solve. This risk is reduced by experimenting at times when life may be assumed to be relatively uneventful in the mill, for example at night.

An inherent drawback of many studies on expert-novice differences is the fact that experts are difficult to recruit. In this study, the same argument applies for all other subjects as well (trainees, novices, and technicians), since there are simply not very many of them available. Since subjects are so difficult to find, special care has been taken to ensure that no subjects were "lost" due to a bad or unsuccessful pilot-experiment, or because the subject becomes, for whatever reason, unwilling to participate any longer.

Some of the studies on the development of expertise are based on very few subjects, sometimes just one (Chase & Ericsson, 1981). Since most of these are exploratory in nature, or only aimed at demonstrating one particular aspect of expertise (e.g. superior memory performance due to chunking), this is not necessarily to be considered a problem. Moreover, it many cases it turns out to be possible to collect many datapoints from one subject. However, there certainly are individual differences between experts, not only in what they know declaratively, but also in the strategies they use to solve problems. Therefore, the approach taken in this study is to investigate diagnostic skill with as many subjects as possible, and thus allowing a wide range of behavior. As a result, this study is partly exploratory but it also provides tests of several hypotheses, as described in the subsequent chapters.

Different techniques are used for the investigation of different types of knowledge. The experiments geared towards an investigation of system knowledge make use of highly structured interviews, drawings by the subjects and variants of thinking-aloud protocols. The experiments geared towards an investigation of diagnostic strategies use thinking-aloud protocols, followed by a protocol analysis. These protocols provide a much more detailed description of the cognitive processes than just the final answer a subject gives. The use of protocols enables inferences about intermediate steps such as subgoals, and attention to specific aspects of the problem. Protocol statements are not treated as introspective descriptions of psychological processes, but rather as overt reports of mental activity the subject is aware of, but usually fails to verbalize. For an extensive overview of the technique of protocol analysis, the reader is referred to Ericsson and Simon (1984).

1.4 Outline of the study

The study can be outlined as follows. First, in chapter Two a description is given of paper making, to make the reader somewhat acquainted with the domain of study. This is followed in chapter Three by a theoretical analysis of models for diagnosis, together with a model developed as analysis tool and interpretation scheme for the data of the experiments to be reported later on. Chapter Four provides us with first insights in the use and organization of one aspect of the declarative knowledge base: knowledge about possible faults for paper breaks. Chapter Five contains a test of the model for diagnosis as described in chapter Three. In chapter Six a replication of this study with different subjects, different alarm situations, and in a different part of the paper mill is reported. Chapter Seven deals more thoroughly with the question of the availability of system knowledge and its use in problem solving situations. Finally, chapter Eight provides a general discussion of the findings, together with implications for practical situations.

Chapter 2

The task of the operator in a paper mill

2.1 Introduction

This chapter will be concerned with the tasks of operators in general, and more specifically, the tasks of operators in paper mills, and especially in the paper mill studied. The chapter is structured as follows. First, a general outline is given of the task of operators in continuous flow plants, such as paper mills. Second, a description of the paper and board industry and the domain of paper making will be given. Finally, the paper mill whose operators participated in the study will be described.

2.2 The task of process operator

The process operators' work in a paper mill, an example of a typical continuous-flow plant, can be classified into four broad categories as described by Crossman and Cooke (1974):

1. **Control**. The operator must monitor the various gauges, attend to the signs coming from the plant itself, such as noises and vibration, and occasionally carry out special tasks on the product. According to his or her interpretation of these indications, an adjustment is made to the controls when necessary so as to keep the product within specification, and to correct any chance disturbance or drift. Most processes require several more or less independent variables to be controlled simultaneously, so the operator usually has to divide his or her attention between several activities, and generally carries out a regular patrol of the various indicators, making adjustments where necessary.

   The components of the control task may vary widely in difficulty. At one end of the scale is the simple task of keeping a single variable at a desired value by means of direct control (e.g. maintaining a flow by opening or closing a valve.)

At the other end, an operator may have to maintain a combination of qualities in the product by a complex balance of conflicting requirements, which is often the case in the paper making process.

In many plants, the product is changed from time to time without stopping the process, which requires the operator to quickly readjust the process to the new specification so as to waste as little raw material as possible in a substandard product.

Another aspect of control concerns the lookout for early signs of uncoming disturbances, so that preventive actions can be taken. An acute operator may save large amounts of material and money by this means alone. One of the operators observed in this study once said: "If one only starts troubleshooting once the problem appears, then you're too late. Before the trouble occurs, you should know where it will be."

2. **Special procedures and drills.** There are usually set sequences of manipulations to be carried out when starting or shutting down the plant or in particular emergencies. They are often rapid and complicated manual operations, interspersed with some control activity. Training of these procedures may be of considerable duration before these procedures become part of the operators' skills, partly because they will not be used very often and, due to shift work (five shifts), the chances for one particular operator having to carry out these procedures are not very high.

3. **Routine maintenance and cleaning.** This may include routine cleaning activities, keeping the machine stocked with secondary materials, and cleaning the inside of vessels when they become clogged. It may also include certain duties during scheduled maintenance activities and assisting technicians.

4. **Recording and reporting.** The readings of important indicators and gauges, control settings, and the results of special measurements are logged at regular intervals, and disturbances or changed conditions are noted when they occur.

   Apart from keeping written records, operators must also pass on information to their colleagues and to the management by word of mouth. This is important in order to coordinate the operation of the various sections of a large plant.

Apart from these four general categories that play a role in every operator's duties, two more can be added that occasionally are called for:

5. **Supervising.** It may happen that one of the operators acts as a supervisor over his or her colleagues and then is responsible for the operation of a certain part of the installation. In many paper mills, the paper maker acts as a supervisor, and apart from specific responsibilities such as controlling part of the installation, the paper maker is also responsible for the operation of the machine as a whole, which is carried out by delegating certain tasks to other operators (for example the dryer man and the assistants).

6. **Training.** In many plants the training on the job of young operators is carried out under supervision of the chief operator on duty. There appear to be huge differences in the way this part of the task is taken: some operators consider the young trainee as a pair of "extra hands" that can be used in difficult situations for routine aspects, while other operators make a big effort in really training the person for the job: explaining as many things as they can, and giving the trainees as much responsibility as they can handle, and thus enabling them to train the necessary control and troubleshooting skills.

If the work in continuous-flow plants is compared to batch or mass production factories the following differences can be noted:

- In continuous flow plants there is no definite work cycle: setting aside some exceptions as regular recording, the operator acts as and when he judges it to be necessary. In many cases he is free to move about the plant most of the time, but should never leave it unattended; therefore, he needs a relief for meals, and in case of illness;
- On the other hand, there are occasional periods of very intense effort, for instance during startup and shutdown, or when malfunctioning has to be diagnosed and repaired;

- The operators in continuous flow plants have more contact with the technical staff and managers. They are more often asked for information about their plant, and are treated more as team members than operators in mass-production factories;
- Shift work (three, four or five shifts) is very common because of the high capital costs of plant and/or waste of materials involved in shutting it down. This means more responsibility for the operator on evening and night shifts when there are fewer staff members on call;
- Financial incentive schemes based on work measurement (piecework) are rarely applied in continuous flow plants, whereas they are quite common in other industries.

Plants can be controlled in several ways as described by Kragt (1983):

- **Local manual control**, in which the operator monitors and controls the process manually on the spot. He or she has to read a number of indicators (which may also contain a visual observation of the product) and manipulate valves by means of handwheels. This is the very traditional situation of process control.

- **Local automatic control**, in which controlling devices make automatic control possible. One of the tasks of the operator in this situation is to check the values of the process variables, as well as to check the functioning of the instruments themselves. Every once in a while, a switch is made to manual control. In this situation the operator is responsible for a larger part of the plant than in the situation of local manual control.

- **Remote and centralized control**. With the introduction of electric signal transmitters local panels could be combined in central control rooms, with lots of individual indicators, recorders, controllers and switches. Operators working in this situation often have to do a lot of coordination between people in- and outside the control room.

- **Distributed instrumentation**. The interface to the process is turned into a (number of) Visual-Display Unit Computer screen(s). The main reasons for this change in interface design are flexibility and process-control quality. The operator keeps control over the process by means of keyboard-operated units. One of the characteristic features of such a system is the integrated presentation of data. In this situation, the task of an operator moves from direct control of part of the process to monitoring and troubleshooting when necessary.

The introduction of automation has greatly affected not only the productivity and methods of organization, but also the demands made on the job of individual operators. A trend can be noticed from direct manipulation of the process to supervision, and from craftswork to skills based on knowledge of the plant and its equipment. The job of operator turns into a "controller" instead of a "producer", or, as some believe in the paper industry: the traditional paper making skills disappear, and only general operator skills, not particularly related to paper making, are sufficient.

However, as put forward by Bainbridge (1987), by taking away the easy parts of the task, automation can make the difficult parts of a human operator's task more demanding. In an automated situation, the operator is still expected to carry out two tasks: monitoring the process and taking diagnostic actions when necessary. The problem with monitoring arises when an operator is supposed to take over. Several studies indicate that experienced operators with good manual control skills make a minimum of actions and the process output moves quickly and smoothly to the desired level (Edwards & Lees, 1974). These operators show **open-loop** control behavior. In contrast, novices tend to wait for feedback after each corrective action, thus setting the process in oscillation. These novices show **closed-loop** behavior. However, manual control skills deteriorate quickly when insufficiently practiced. Therefore, formerly experienced operators with good manual control skills, having dealt with an automated system for some period of time, will turn into novices in this respect, and one may wonder whether that is sufficiently known by them. On the other hand, the question may be posed how useful manual skills are in a situation where the speed of the manufacturing process is too high to apply manual control skills at a regular basis.

Diagnostic skill at least partly depends upon a thorough knowledge about how the plant functions and how to control it. However, in order to develop effective strategies for diagnosing problems and to be able to build up well integrated knowledge, one needs experience and practice which only develops when one encounters these problems. In plants that have been automated to a large extent, many small disturbances of process values are handled by the system itself. Therefore, only more serious problems remain.

Due to the reason mentioned above, operators may lack the necessary experience to solve these problems. In a situation of manual control, operators have much better opportunities to find out effective strategies since they spend much more time on parts of the process itself. In the new situation, operators are mainly classroom, or control room instructed, and spend less time developing practical skills. As noted by Bainbridge (1987), there is some concern that the former manual operators, who are now monitoring the automated systems, are riding on their previously acquired skills, which later generations of operators cannot be expected to have, unless the training systems change. The same comment is sometimes made elsewhere, for example the management of the paper mill of this study hold it as a common opinion.

2.3 The paper and board industry

According to recent predictions, the paper industry will continue to be a "growing industry" at least until 1995 (VNP, 1989). In 1988, the paper consumption in the world showed an increase of 4.7% to 224 million tons, compared to 1987. The year 1989 was the seventh year in a row in which sales of the paper mills were increased substantially. To establish those high sales and to meet the corresponding high demands, it is considered important to keep constant high levels of investments, both in new machines and in new technology. The level of investment each year is on average more than 12% of the annual turnover, and in 1986, 1987, and 1988, in all sectors of the Dutch paper and board industry there have been investments in new machines or modernizations of older ones. The Dutch paper and board industry produced in 1988 2.475.000 tons with 70 machines, localized in 32 paper mills with 8300 jobs. In the European Community as a whole in 1988 there were 952 paper and board mills with 170.500 jobs. Very important for the development of this branche is the continuously increasing use of waste paper as base material in paper mills. To accomplish the recycling of waste paper, the recollection of it is very important. Therefore it is interesting to note that in The Netherlands 66% of waste paper that can be recycled actually finds its way back to the paper mills. In this respect, The Netherlands scores highest in the world.

2.4 Paper making

There are many types of paper, each being differently produced on specific machines, such as newsprint paper, writing paper, coated paper for illustration purposes, paper for the corrugating industry, wrapping paper, solid board, and tissue paper.

The manufacturing of paper is, regardless the type of machine, always based on the same principle. A mixture of fibers and water (pulp) is led onto a wire screen (for example a fourdrinier wire) so that much of the water drains through the screen and a wet web remains (the wet section or wet end). The web is subjected to vacuum and presses (the press section) to remove more water. To accomplish this effect, press felts are used, principally made from synthetics. Finally the paper is dried by passing it over steam heated cylinders (the dryer section or dry end). It can be finished or polished by threading it through a series of vertically stacked heavy cast iron rollers (calender rolls) interchanged by paper rolls. To obtain pulp of good quality, cleaning processes are very important, to remove all foreign matter such as iron wires, staples and plastic. This type of cleaning is especially important when waste paper is used as main base material. This cleaning process is repeated several times with greater levels of refining at every stage (thickstock and thinstock cleaning, by means of thickstock centrifugal cleaners, fiberizers, screens and several thinstock centrifugal cleaning batteries). In between the several processes there is a number of buffer chests. A schematic overview of the paper making process is given in figure 2.1.

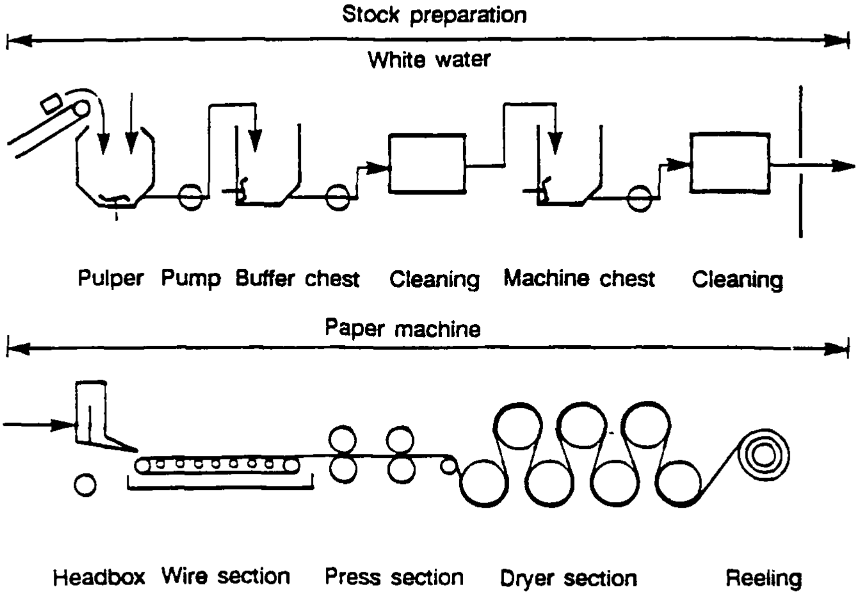


Fig. 2.1 A schematic overview of the stock preparation and paper making process.

Paper furnish or pulp consists of a mixture of purified vegetable fiber and water brought together in predetermined proportions. This suspension is commonly known as stock. The weight of fiber contained in a given volume of stock is called "consistency". The control of consistency is one of the prime requisites of paper making and it is essential that this is done after each stage of the stock preparation process. Swings in pulp consistency that cannot be controlled anymore may cause swings in basis weight, affecting sheet formation, pressing, drying, calendering, wet and dry end paper breaks, winder snap-offs, etc.

Water is heavily used in paper making, and it is therefore not very surprising to notice that many paper mills were initially founded near a river. The amount used by paper and board mills varies widely. More water is used in the manufacturing of paper and board than any other raw material. Besides serving as a carrier of fibers, it is essential to the pulping and paper making operations as a solvent for required chemicals. Nowadays, most, if not all, of the water is reused. This increased reuse of so-called "white water" (water containing fibrous material) has resulted in several economic savings: the fibrous material suspended in the water and added chemicals are saved, environmental pollution is minimized, less water from outside sources, such as rivers, is required, and energy savings (heat!) are possible. The water that cannot be reused goes through a purification process before it is lead back to its source.

Paper making is an energy consuming process, especially since the drying of the paper requires a lot of heat (steam). Therefore, it is considered important to get the paper out of the press section with as little moisture in it as possible. Many companies increasingly try to cover their energy demands through the operation of company owned generators driven by for example gas powered turbines.

Paper sometimes possesses certain unwanted qualities that render it unsuitable for its intended end use. These so-called paper defects can often be prevented by exercising extra care during the whole paper making process, but in other cases they are almost inherent to the process. Examples of paper defects are variation in paper weight, wet streaks, wrinkles, holes, and felt marks. Felt marks may be caused by badly worn felts in the press section, but can also occur when a felt gets older and should be replaced. Since one will postpone the decision to change felts as long as possible, felt marks may be the inevitable result at the end of the life cycle of the felt. Some paper defects may be caused by mechanical defects, such as a badly driven pump, which should be alleviated by maintenance personnel. On the other hand, paper defects may occur due to mistakes on the side of the production personnel.

Another common problem in paper making is the occurrence of paper breaks. Due to some irregularity the paper tears off. This may occur in the wet end or in the dry end of the machine. If the paper tears off in the wet end (before it reaches the dryers), this so called "wet broke" must be removed from the presses before a sufficient quantity has accumulated to overflow and causes damage. Therefore, the broke is lead into the press pit, and the task of the operator is to ensure that the broke pits are working properly, which means leading it back to the stock system for reprocessing.
"Dry broke" is produced at the dryers, calender stack, reel, winder or "tear offs" at the finishing room. There are usually holes in the floor under the calender and reel extending the full width of the sheet, through which the paper goes to the repulping process.

Elaborate descriptions of paper making can be found in many textbooks, such as Calkin and Witham (1957), and MacDonald and Franklin (1970). Appendix C contains a glossary of paper making terminology used throughout this study.

In many paper mills, stock preparation and paper making on the paper machine are separate departments, run by different operators. Stock preparation needs two or three operators, depending on the size and complexity. The paper machine is handled by a crew of four to six operators. The size of a paper machine is about 100-180 meters in length, with a common width of between 2 and 7.5 meters. The speed of paper production varies between 100 meters per minute (board) up until as high as 1500 meters per minute (newsprint).

Almost all paper mills are continuous flow plants, operating in three, four, or five shifts. Not much is known about the selection criteria for operators, but it is assumed that at least nowadays the selection is mainly based on secondary school education and general impression. Operator training in The Netherlands differs in content from mill to mill, and can take several forms, secondary school education not included:

- Theoretical training courses off-the-job of increasing difficulty, which tend to be general operator courses, not geared specifically to the paper industry, and leading to nationally acknowledged certificates (VAPRO-A, VAPRO-B, VAPRO-C). These courses can be taken in special schools, but the certificates can also be obtained through self-study, combined with the requirement of a job as operator;
- Theoretical training geared towards knowledge of the specific processes in the paper industry, but not geared towards the processes in the operator's own mill. These courses are given by VAPA, the Dutch educational institute for the paper and board industry, which is also funded by this industry;
- Theoretical training geared towards knowledge about the specific processes in the operator's "own" paper mill. This type of education is typically given by people from the mill itself, sometimes on the job, sometimes off the job, sometimes both;
- Practical training on the job. The practical training most of the time consists of working with a certain crew of operators. The young operators are expected to work and learn at the same time. However, it strongly depends on their supervisor how much practical experience can be obtained in actually controlling the process. Some supervisors do not delegate any control to the trainee, while others spend quite some effort in training the young operator how to control the installation. The question remains though whether this is sufficient in building up good control and diagnostic skills.

Most of the paper mills use a combination of the above mentioned methods to train their operators.

2.5 Description of the paper mill of the study

The paper mill whose operators participated in this study has about 335 employees. The major part of the production tonnage (approximately 90%) is used as supply for the corrugated case making industry. The mill uses only waste paper as its base material. Three paper machines are operated, which, although varying in capacity, operate roughly on the same basis. They have different degrees of automation though. The first two (oldest) paper machines are controlled by remote and localized control (they have typical "classical" control rooms). The difference between the two is the number of control loops automated, one of them having only a few automatic control loops, the other having many more. The third machine, which has only been built in 1986 and was recently extended, is much more computer controlled and has distributed instrumentation as outlined in section 2.2. The organization of the paper production work moved to five shifts in the early 1980's. Both this change from four shifts to five, coupled with the installation of the new paper machine in 1986, necessitated the hiring of many new operators in this period. The fact that the training of these new operators did not always proceed smoothly, has increased the interest in training issues. In the past, most of the work was coordinated by informal communication: there were rather short lines from the managers to the production workers and maintenance personnel. However, due to the growth, more formal structures are now appearing. Preventive maintenance of the installation is scheduled at regular intervals. However, there are still many occasions that corrective maintenance is needed. This type of maintenance is mostly carried out at times that parts of the installation are shutdown. (Of course, sometimes the paper machine is shutdown because some mechanical defect occurs, necessitating corrective maintenance, but the paper machine can also be shutdown for other reasons, related to the production, such as changing a felt).

Operators have a strong drive towards the production of paper, and they are very much pushed towards making paper. This implies that as soon as some irregularity, such as a paper break, occurs, everybody available (people operating the winder as well) helps in getting the installation to work again. The production of paper is considered the principal goal in the mill.

Expertise in paper making is not concentrated within one person, but is spread over operators, due to shiftwork and an incomplete transfer of information over shifts. Historically, not much formal training was given to the operators, and the existing expertise was mainly acquired on the job. Recently, much more effort is put into more systematic training, and many operators nowadays are striving for one or more of the above mentioned certificates.

The management of this paper mill may be considered ambitious, and is willing to invest in new technology. It positions itself among the leading market-oriented recycling operations serving the packaging industry. The mill is operating well economically. It is fair to say that the atmosphere can be considered very good, and employees are quite prepared to put extra effort into their job when required.

Chapter 3

Models for diagnosis

3.1 Introduction

The task of troubleshooting or diagnosis (the terms will be defined more precisely later on) may be described rather simply. Given that a system is not functioning properly, the troubleshooter must attempt to locate the reason for the malfunction and must then repair or replace the faulty component. Quite a lot of research has been carried out characterizing diagnostic behavior and diagnostic skill, resulting in models aimed at describing and explaining diagnostic behavior and skill. Models for diagnosis have been developed from at least two different viewpoints: artificial intelligence and psychology or human engineering. Characteristic of AI-models is the principal interest in building intelligent machines, able to carry out tasks at a certain, preferably high, level of performance. These models do not necessarily have a close relationship with human behavior, and may not take human intelligence as a starting point. Psychological models on the other hand, aim at describing and explaining human behavior, and these models may use artificial intelligence techniques as modelling tool, but are not necessarily interested in building an intelligent machine. In the case of diagnosis, both approaches have resulted in interesting models. A number of those models will be discussed in this chapter. As examples from AI-approaches to diagnosis the relevant pieces of KADS (Knowledge Acquisition and Documentation Structuring), as developed by Breuker and Wielinga, and approaches to model-based reasoning (De Kleer, Davis, Hamscher, and Bakker, among others) will be discussed. As representative of psychological model building with respect to diagnosis in technical systems, work by Rasmussen and Rouse will be discussed. This chapter will conclude with a proposal for a general model for diagnosis aimed at describing and explaining diagnostic task strategies of expert and novice operators. Before this discussion starts it is necessary to define what is meant by diagnosis or troubleshooting, since the terms are used differently in the literature. Sometimes diagnosis is only defined as the process from identification of the symptom to the determination of the fault. In other cases, especially when one speaks about troubleshooting, the whole process of symptom identification, fault determination, and compensatory actions is taken into consideration. In this study, both terms ,"diagnosis" and "troubleshooting", will be used, both meaning the complete process from the identification of symptoms to the taking of appropriate corrective actions.

3.2 AI-approaches to diagnosis

3.2.1 KADS

One of the methodologies for expert system development that puts a lot of emphasis on modelling expertise is the KADS (Knowledge Acquisition and Documentation Structuring) methodology, developed by Breuker and Wielinga (Wielinga & Breuker, 1984; Breuker & Wielinga, 1985). The KADS methodology is a knowledge acquisition methodology aimed at offering a more structured approach to building knowledge-based systems than the often used experimental approach of prototyping, in which systems are built incrementally and knowledge acquisition goes on continuously.

According to the KADS methodology, knowledge of experts can be analyzed and described at different layers. The first layer, the *domain layer*, contains static knowledge about the domain: domain concepts, relations and complex structures such as models of processes or devices. Knowledge at this level is supposed to be *task neutral*: it is not represented in a special way in order to support a particular task. It is questionable, though, whether domain concepts are really independent of a supporting task. Previous research (Schaafstal, 1989) showed that, dependent on job structure, people use very different domain concepts as primitive terms to reason with. For example, the concept "pump" has a different meaning for people with different jobs. An electrician means by pump the *engine* of the pump, since the engine of the pump is the only part of the pump he is concerned with in his daily routines. An operator, concerned with operating the device, uses pump in a more global sense: a pump is a device used in transportation of stock from one place to the other.

The second layer of expertise is the *inference layer*. In this layer the inferences that can be made based on the knowledge in the domain layer are described. At the inference layer two types of entities are described: meta-classes and knowledge sources. Meta-classes describe the role that domain concepts can play in a reasoning process, for example in diagnosis: a certain concept may be a hypothesis or a remedy. Knowledge sources describe what types of inferences can be made based on the relations in the domain layer. Knowledge sources are defined at a very general level. An example from diagnosis is "compare" by means of which is observed whether a difference exists between the observed and expected output. Another example is "test candidate", which determines whether a certain candidate is the real cause of the complaint or that the candidate is in fact working correctly. At the third layer, the *task layer*, knowledge about how goals and tasks are reached is specified, given the specification of inferences at the inference layer, resulting in a task structure. The task structure of a diagnostic process in which reasoning with empirical associations plays a central role may look like the following:

- Try to associate a complaint (temperature of 40° C) with a certain hypothesis.
- Abstract from complaint to a more abstract complaint (high fever).
- Associate from abstract complaint (high fever) to hypothesis (influenza).

The fourth layer is the *strategic layer* in which knowledge resides which allows a system to make plans, control and monitor the execution of tasks, diagnose when something goes wrong and find repairs for impasses. This fourth layer allows for flexible behavior, in the sense that it contains knowledge to change execution of plans triggered by certain circumstances. An example from diagnosis: in normal circumstances, people will try to diagnose what the cause is of a certain complaint in order to prescribe the correct medicine. However, when someone bloods heavily, we will first apply some remedy to stop the bleeding, before a diagnostic process starts. Thus, a problem requiring immediate action will change the execution of the normal diagnostic routines.

The four layers of expertise and their relationship to each other are summarized in table 3.1.

Table 3.1 *The four layers of expertise in the KADS methodology.*

Layer	*Elements*	*Relationship with other layers*
Domain layer	Concepts/ Relationships	
Inference layer	Knowledge sources Meta-classes	Description of domain layer
Task layer	Tasks/Goals	Applies to inference layer
Strategic layer	Strategies/ Meta-knowledge	Controls task layer

Within the KADS methodology a distinction is made between *conceptual models* and *interpretation models.* A conceptual model is a high-level specification of the expertise required to build a knowledge-based system, which is used for both design and implementation. Thus, a conceptual model contains both task and domain information. However, as discussed in Breuker and Wielinga (1987), experts often use systematic problem-solving methods that are domain-independent. Pople (1982) claims that an important aspect of clinical reasoning is the structuring of diagnosis on the basis of meta-level knowledge. The same conclusion is drawn by Stefik (1981) on planning. Konst, Wielinga, Elshout, and Jansweijer (1983) found similarly structured behavior in skilled problem solvers in physics problems. Also, analysis of diagnostic systems has revealed task structures that are typical for diagnosis but are relatively domain-independent (Bennett, 1985; Clancey, 1985; Szolovits, 1981). This implies that expert behavior that is seemingly domain-specific may originate from higher level problem-solving methods, that are well structured and have some degree of domain independence. Domains may contain different concepts but they may share ways of using knowledge at a higher level of abstraction. Therefore, a conceptual model for a certain task can be abstracted from its domain specific content and thus function as a skeletal model for the conceptual model of expertise for a similar task in a different domain. This abstracted task is called an interpretation model. A description of an interpretation model contains a discussion of the inference structure, the task structure and the strategy level. Within the KADS methodology a number of interpretation models has been developed for different tasks, such as diagnosis, monitoring, planning and design. Since this chapter is concerned with an overview of different models for diagnosis, only the interpretation models for systematic diagnosis and heuristic classification will be discussed. Systematic diagnosis is diagnosis based on some kind of model of the device, which is used to reason about the causes of the complaint. Heuristic classification on the other hand uses heuristic rules to obtain the likely causes for a complaint. These heuristic rules are empirical rules, in which symptoms are associated with possible causes. The discussion here will mainly concentrate on meta-classes and task structures used within those models. For a more detailed overview of interpretation models, see Breuker, Wielinga, van Someren, de Hoog, Schreiber, de Greef, Bredeweg, Wielemaker, Billault, Davoodi, and Hayward, (1987).

Systematic diagnosis

Interpretation models for systematic diagnosis can be divided into two categories: diagnosis by localization and causal tracing. Diagnosis by localization can be viewed as diagnosis in the sense of finding the component that is causing the complaint. It is possible when:

- The possibility of developing a part-of model exists, a model in which several components can be distinguished;
- It is possible to test components independently from the rest of the system;
- Detailed specifications of the system behavior, or output values of components are available, or can be obtained.

Typical domains for diagnosis by localization are troubleshooting devices, such as computers, in which several components can be tested individually and in which the design outline may serve as a starting point for development of the part-of model. The reader should notice that diagnosis by localization is similar to the topographic search strategy defined by Rasmussen (1986), which will be discussed later on in this chapter.

Diagnosis by causal tracing can be viewed as finding the cause for the complaint, which is not necessarily identical to finding one (or more) faulty component, since it may also be a fault in a certain process (for example, insufficient steam capacity). It is possible in the following situations:

- It should be possible to explain the function of a system by causal relations, for example the mechanical-causal relations of the components of an engine;
- It should be possible to obtain data on a significant part of the states implied by the model.

Diagnosis by causal tracing can typically be found in troubleshooting mechanical systems, such as engines. Often, both causal tracing and localization play a role, since one point of view may not be sufficient to systematically eliminate hypotheses.

The meta-classes and their types of domain concepts for localization, respectively causal tracing are shown in the next table.

Table 3.2 *Meta-classes and their types of domain concepts for localization, respectively causal tracing.*

meta-class	*localization*	*causal tracing*
system model	part-of model	causal model
complaint	faulty system	faulty state
universe of observables	observable output variables	observable states
hypothesis	(sub)system containing faulty component	sub-network
variable value	observed output value	observed state
norm	output specification	top state causal subnetwork
difference	faulty component	faulty state
conclusion	faulty component	cause of complaint

The meta-classes describe the role that can be assigned to the various domain concepts in the reasoning process. Dependent on whether we are concerned with diagnosis by localization or diagnosis by causal tracing, different aspects of the domain concepts are used. Since localization is concerned with finding the system component causing the complaint, the terminology used will be in terms of systems, and system components with its outputs. Causal tracing is aimed at finding the right cause for the complaint, which may well be a wrong process (for instance, insufficient steam used, causing too wet paper). Thus, the system model used to reason with in localization is a part-of model, such as a hierarchical representation of a device with all its components and subcomponents. In causal tracing, a model would be used consisting of all the causal relationships in the device (pump 2 causes stock to flow from chest 2 to chest 3). The complaint might be a broken pump in localization, and no stock flow from chest 2 to chest 3 in causal tracing. The universe of observables (things that can be checked) consists in localization of output variables of a certain component (for instance, amperage used by a pump). In causal tracing, it might be something like a state of the system. The hypothesis formulated in the two situations are also somewhat different: in localization it would point to a (sub)system or a faulty component (something that can be touched), while in causal tracing it might be something as "the heat transmission system", a sub-network of causal relationships. The norm should be viewed as the specification of correct output values of systems or correct causal relationships. The meta-classes "difference" and "conclusion" speak for themselves. Obviously, the models have a relationship to each other, since although a system may be described in terms of the causal relationships between the various functions, in the end these causal relationships are implemented by components or sets of components.

Both submodels for systematic diagnosis, localization and causal tracing, have the same inference structure which is shown in figure 3.1.

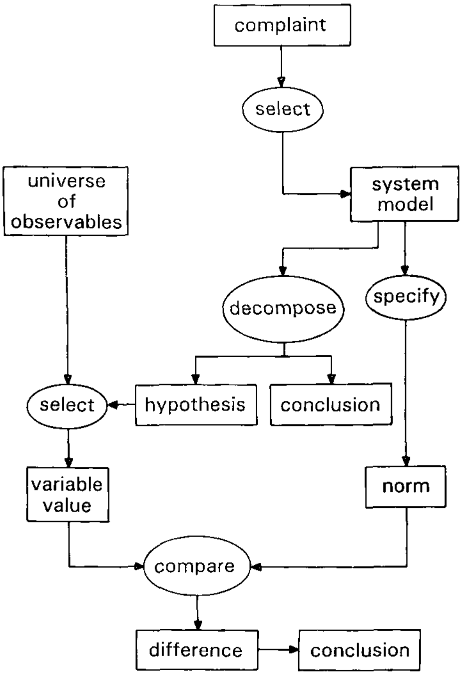


Fig. 3.1 Inference structure of systematic diagnosis.

After formulation of a complaint, the diagnosis starts with the selection of a system model, which can either be a consist-of hierarchy of components (in localization) or a causal net (in causal tracing). The model is decomposed to be able to formulate a hypothesis about likely causes for the complaint. If further decomposition is impossible, a conclusion about the cause for the complaint is reached. To test the hypothesis, a variable value is selected from the universe of observables of the output of the hypothesized faulty component. This value is compared to a norm about the expected normal output value or function. If a difference is found between the norm and observed value, the reasoning process stops and a conclusion is reached.

Systematic diagnosis has the following task structure:

```
task
        find(diagnosis)
                select(system model)
                        WHILE(no conclusion)
                                decompose(system model)
                                        WHILE(number of hypotheses in differential (= set
                                        of hypotheses given the complaint)) > 1)
                                                select(variable value)
                                                specify(norm)
                                                compare(variable with norm)
```

Diagnosis starts with selection of a system model. After a system model is selected, a number of hypotheses, called a "hypothesis differential", is generated by the decomposition task. The hypotheses in the differential are tested by comparing the variable value with the norm. If a difference is found, all other hypotheses in the differential are rejected and the system model is decomposed again. This process is repeated until the lowest level of decomposition is reached. The remaining hypothesis is confirmed and a conclusion is reached.

Compared to heuristic classification, which will be discussed in the next pages, systematic diagnosis has a more algorithmic reasoning structure, since hypotheses can be eliminated without using likelihoods of occurrence. Furthermore, systematic and heuristic diagnosis do not necessarily have to be separate tasks, but may be combined. For example, heuristic associations may be used first to focus on likely malfunctions, and localization and causal tracing may play a role as back-ups, supporting additional evidence, or for justification purposes. For example, Steels and van de Velde (1985), summarized in Breuker et al. (1987) show how heuristic shortcuts may emerge from reasoning with a causal model.

Diagnosis by heuristic classification

According to Clancey (1985), diagnosis in knowledge based systems is solved by heuristic classification. The inference structure of heuristic classification is shown in figure 3.2.

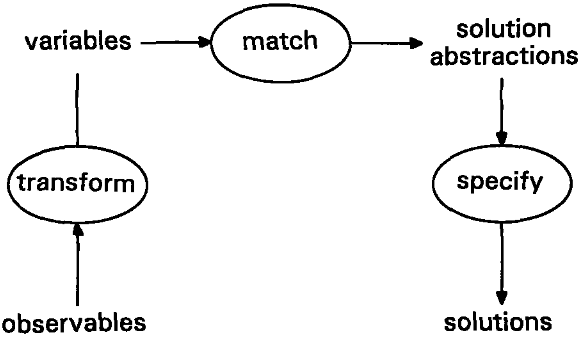


Fig. 3.2 Inference structure for heuristic classification.

This figure should be read in the following way. An observable, for example a temperature of 40° C, is transformed into the variable high fever. This variable matches some general descriptions of solutions, or causes for the disease, for example bacterial infection. This solution abstraction can be further specified into more concrete solutions, such as a certain type of bacterial infection.

Heuristic classification can be either data driven (forward reasoning) or goal-driven (backward reasoning). It has the following task structure:

Data driven diagnosis

obtain data
transform(variables)
match(abstracted data)
specify(solution abstractions)

Goal-driven diagnosis

specify(solution abstractions)
match(abstracted data)
transform(variables)
obtain data

Now we have discussed the interpretation models for diagnosis it is worthwhile to make some comments. One of the things that can be brought to bear about the usefulness of KADS is that in real-life situations most tasks do not consist solely of just one generic task, but will be a composite of several generic tasks. For example: diagnosis, in the sense of fault finding, and the undertaking of appropriate corrective actions to remedy the problem, as stated in section 3.1, both play a role in troubleshooting paper machine problems. Therefore, when modelling expertise in "real-life" situations one has to be aware that the model used to interpret human behavior may have to be a composite of several generic models. The question may be posed whether it is possible to describe characteristic task strategies in real-life domains at a higher aggregation level than the KADS interpretation models. This will provide us with a better understanding of those tasks and their interaction. As long as the tasks are still rather general in nature, the results can be generalized to similar tasks, and therefore, the models are also theoretically interesting. For example in diagnosis, it should be possible to formulate a model for diagnosis including both fault finding, and undertaking of repairs and the interaction between both tasks, as a better guideline for knowledge structuring processes.

However, the KADS methodology offers a starting point from which such a real-life model can be developed, although the interpretation models are rather sketchy in nature. KADS interpretation models are in fact competence models of behavior, with the behavior of experts in a certain domain only used as a starting point. The study reported here aims at describing a psychological performance model for diagnostic strategies in a technical domain, in which attention is given also to the role of underlying knowledge. When heuristics fail, or are not present, what knowledge about the installation or the production process is used by operators to reason about the posed problem? In this respect, not much help is offered by the KADS methodology, since the system knowledge which is used by an operator to reason with is part of the first, static layer of expertise, and is domain-dependent. The question is however, whether it is possible to describe domain models at a higher aggregation level than the domain itself, thus enabling a more abstract and general description of used knowledge. For example, certain types of models may play a role in several domains, and its characteristics may be formulated in a more abstract manner. In this respect there are at least two types of models with these characteristics described in the literature, models based on the *structure* of the device (Bakker, Hogenhuis, van Soest, & Mars, 1988; Davis & Hamscher, 1988, among others), and more *functionally oriented* models (for example: Benjamins and Jansweijer, 1989), which will be discussed in the following part of this section.

In conclusion: KADS offers a strategy model at a very general level for diagnosis, which certainly should be adapted for any real-life task. KADS is not concerned with the interaction between strategy and underlying knowledge, apart from a change in applied strategy from heuristic reasoning to systematic diagnosis.

3.2.2 Diagnosis based on structure and behavior

In diagnosis in technical domains it is not always necessary to derive a model of diagnostic reasoning exclusively from experiential knowledge. In these domains there often exists knowledge about the underlying theory of the device. Derived from the design of the device, there is a description of the intended behavior and structure of the system. Based on the design, a specification can be given of the components that are present in the system and the way they are interconnected. One of the main ideas of model-based diagnosis is that as much of the available information as possible about the device and its components should be taken into account. Models have been developed for several domains, such as circuit schematics with resistors and diodes (for example, de Kleer, 1976; Brown, 1976; Dague, Raiman, & Deves 1987), circuit schematics using logic gates and higher-level digital components (for example, Davis, 1984; Genesereth, 1984), piping and instrumentation diagrams, including components such as valves, potentiometers and lamps (Scarl, 1985), and models of human physiology (Kuipers & Kassirer, 1984; Long, Naimi, Criscitiello, & Jayes, 1986).

There are a number of important assumptions with respect to model-based reasoning, in which some underlying theory about structure and behavior of the device is taken into account (Bakker & Mars (1989)). Diagnosis is triggered by the observation of anomalous behavior. This behavior is assumed to be caused by a change in the system, for example a defective component. Diagnosis aims at finding the defective components in the system that changed the behavior. When those changes are incorporated in the model of the correctly working system, a model of the defective system will be obtained. Apart from observations, a model of the system is needed, containing a description of its structure and the behavior of its correctly working components. In summary, model-based diagnostic reasoning using structural descriptions of a system has a number of prerequisites (Bakker, Hogenhuis, van Soest, & Mars (1988); Bakker et al. (1989); van Soest, Bakker, Hogenhuis & Mars (1988)):

- **A description of the structure of the device**

  This description consists of a list of the components of the device, and their (directed) connections. Since the model is aimed at diagnosis based on structure, the functioning of components is not included.

- **Observations**

  Only observations have to be specified at which point a discrepancy has been observed between the actual and expected behavior.

- **A qualitative model of the behavior of the components in the device**

  By a qualitative model of the behavior of the components in the device a description is meant of how a component propagates a false (f) or a correct (c) on its input to its output, depending on its state: working (w) or defective (d).

How propagation takes place is specified in fault models. A weak fault model, as depicted in table 3.3, starts from very weak assumptions: a working component receiving correct inputs (c) will have a correct output. In all other situations, prediction is impossible: a faulty input (f) may result in either a correct or faulty output, and a defective component (d) may cause a faulty output, but that is not necessary.

Table 3.3 *The weak fault model.*

Output	Component	
Input	w	d
c	c	c/f
f	c/f	c/f

Table 3.4 *The strong fault model.*

Output	Component	
Input	w	d
c	c	f
f	f	f

The strong fault model, as shown in table 3.4, assumes that only when both the input and component are correct, the output will be correct. In all other cases, the output will be false. In between the weak and the strong fault model are intermediate fault models, of which one is depicted in table 3.5. If the component is working and the input is correct, the output will be correct. If either the component works, but the input is false, or the input is correct but the component is defective, a false output occurs. However, a faulty input in a defective component leaves us with an undecidable situation.

Table 3.5 *The intermediate fault model*

Output	Component	
Input	w	d
c	c	f
f	f	c/f

Generally speaking, it is possible to use more than one fault model to describe all components of a device.

In the literature, two reasoning methods have been described for model-based diagnosis (Davis (1984); De Kleer and Williams (1986, 1987)). Since this study will mainly concentrate on the nature of the models used, these reasoning methods will only be discussed briefly. In the method developed by Davis (1984), firstly candidates are generated based on observations of incorrect outputs. These candidates are generated by the collection of predecessors of the incorrect output. By means of constraint suspension, the elimination of constraints belonging to a candidate, the different candidates are tested whether they are consistent or not. Checking of consistency happens by forward and backward propagation. If the output of a component is inferred to have different values, an inconsistency appears, and the candidate is eliminated. In the example of figure 3.3 testing of candidate [M2] leads to an inconsistency. Based on the assumption that M1 and A1 are functioning correctly, and due to the observations, the output of M2 should be 4. However, assuming M3 and A2 function correctly, the output of M2 should be 6. Thus, a defective M2 cannot explain the observations and therefore M2 is eliminated as a candidate.

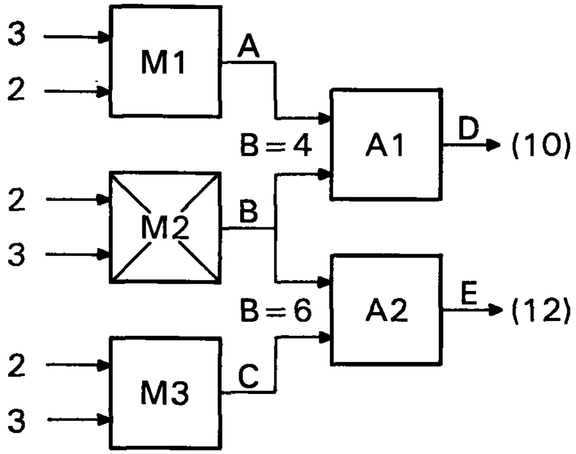


Fig. 3.3 An example model. M1, M2 and M3 are multipliers, A1 and A2 are adders.

The method developed by De Kleer and Williams (1986, 1987), the General Diagnostic Engine (GDE), is based on a different principle. Important in this method is the concept of *possible diagnosis*, which is a hypothesis about the way the faulty system deviates from the model of the correctly working device.

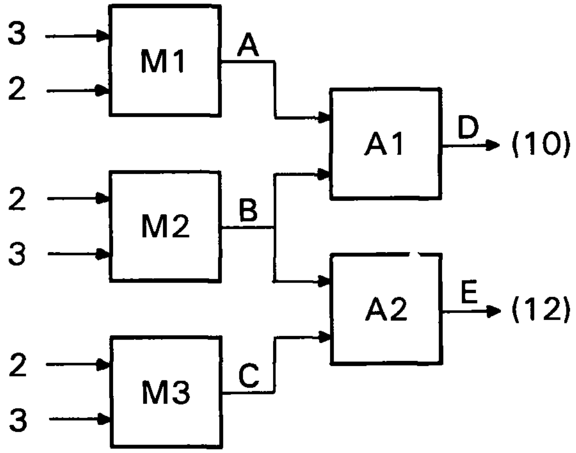


Fig. 3.4 An example model. M1, M2 and M3 are multipliers, A1 and A2 are adders.

In figure 3.4, A1 is a possible diagnosis, since a defective component A1 can explain the observations. It is not always easy, and sometimes even impossible, to determine all possible diagnoses, because of the exponentially growing number of possibilities with the number of components in the device. Therefore, possible diagnoses are determined by *conflicts*, a number of assumptions about the system that results in conflicting statements. For example, in figure 3.4, if it is assumed that M1, M2, and A1 function correctly, then the output of A1 should be 12. However, the observed output of A1 is 10. Therefore, the assumption that M1, M2 and A1 function correctly leads to a contradiction. These three components generate a conflict: <M1, M2, A1>. GDE describes a method to generate diagnoses from conflicts.

The principal aim of model-based diagnosis is to build intelligent systems, able to carry out the diagnostic task in a quick and efficient manner. These models are not meant to be psychologically plausible accounts of diagnostic behavior.

The question remains whether, psychologically speaking, evidence can be found for humans taking structural models of devices into account in their diagnostic reasoning process. In this area, an interesting study has been carried out by Jansweijer and co-workers (Jansweijer, Benjamins, and Bredeweg, 1989). They studied service engineers while diagnosing complex devices, such as an UV-recorder. The results showed that the primary model used by service engineers is a model of the function of the device on different levels of abstraction. The service engineer knows the set of subfunctions that realizes a higher level function. Each subfunction can itself be described in the same way, until, at the lowest level, a function is realized by a component that cannot be decomposed further, or of which the decomposition is of no interest anymore. Thus, on a low level, the model of the service engineer corresponds to a static description of all of the components and the relations between the components. However, Jansweijer et al. (1989) concluded that this level is seldom of interest to the service engineer, since in standard repair practice, complete assemblies of components will be replaced, rather than the individual components. This implies that at least in this domain, human problem solving uses several types of models, functional and structural, together with experiential, heuristic knowledge in which a direct relation exists between observed phenomena to the causes of these phenomena or even to their repairs. Studies of problem solving suggest that the development of competent problem solving is associated with the development of strong methods: the heuristics mentioned above (see for example Anderson, 1983). This implies that presumably both good models of devices, represented declaratively, and strong problem-solving methods will constitute human expertise. Hamscher (1988), who developed a structural model for troubleshooting digital circuit boards, admits that his model performs roughly at the level of the human novice. Thus, although very useful, and sometimes even necessary when no human expertise is available, or too costly, structural models alone cannot yet play a large role in designing intelligent systems, although in combination with other models and reasoning methods, they may certainly improve on the existing expert systems that are based only on shallow, experiential knowledge. It is certainly interesting, though, to investigate what role structural models play in a complex domain such as troubleshooting in a paper mill, in the same line as the study by Jansweijer et al. (1989). Therefore, in the following chapters, data pertaining to this topic will be discussed when appropriate.

3.2.3 Conclusions

Summarizing the two AI-approaches towards diagnosis discussed in this chapter, it can be concluded that KADS on the one hand offers interpretation models for diagnosis, but does not have a great deal to say about the nature of models used in diagnostic reasoning when, for example, heuristics fail. Also, KADS interpretation models are no real-life models, since in real-life tasks many tasks are often combined. For example: real-life diagnosis may comprise certain monitoring and compensatory aspects as well. Finally, KADS models should be regarded competence models in which no close relationship exists with human performance. Diagnostic systems based on a model of the structure of a device are not yet able to perform above the level of a novice, but may offer a starting point for thinking about the nature of models in human expertise. Combined with models of heuristic knowledge they seem rather promising for the future of so-called "second generation expert systems", in which heuristic knowledge is combined with a deeper model of the device.

Both KADS and approaches to model-based reasoning are real AI-models in the sense that they do not take human intelligent behavior into account as starting point for their model, and there are no explicit considerations with psychological validity. However, they do offer starting points for the development of a model for human strategies applied in diagnosis. Also, they provide hints with respect to the kinds of models of domain knowledge one has to look for in human problem solving behavior.

In the literature, there have also been various accounts of psychological models of diagnostic behavior, not aimed at describing the optimal solution methods for a given problem, but geared towards an adequate description and explanation of human behavior, as a first step towards better training methods, support systems and interfaces. Two of these approaches, those by Rasmussen and Rouse, will be described in the next section.

3.3 Psychological approaches to diagnosis

3.3.1 Rasmussen

Rasmussen's basic study of diagnostic strategies was carried out in an electronic instrument repair shop under real-life conditions (Rasmussen & Jensen, 1974), using interviews and verbal protocols as research tools. The results from this study have later been generalized to diagnosis in computer systems and process plants, together with an analysis of error reports from power plants. The most important findings of the study on technicians in the electronic repair shop were the following, as described in Rasmussen (1986):

- The trained technician, contrary to the designer, used many observations in a sequence of simple decisions. He uses a general search procedure that is not dependent on the actual system or specific fault. He treats the observations individually in a stream of good/bad judgments that is informationally uneconomic, since no previously experienced faults and disturbances are taken into account and only good/bad judgments are made. However, these observations are made very fast, and thus, it may pay off to make a few more observations;
- The technician defines his task primarily as a search to find where the faulty component is located in the system. He is not concerned with explanations why the system has the observed faulty response or understanding the actual functioning of the failed system;
- Search procedures are organized as a search through a system that is viewed as a hierarchy of units and subunits;
- The general structure of the search can be broken down into a sequence of three different search routines, which are used to identify the appropriate subsystem, stage, or component. These three search strategies employed by the technicians are *functional search*, *topographic search*, and *search by evaluation*. In the functional search, the topographic reference is obtained from the normal functional relationship between a feature in the system response and a specific part of the system. This search is a special type of the topographic search as discussed below. The information pattern is scanned and familiar features are judged individually in a stream of good/bad judgments. If a response feature is judged faulty, attention is typically turned immediately toward the subsystem related to that function. In topographic search reference to the location of the fault is obtained from the topographic location of a measuring point. Search by evaluation of the fault is used when the technician derives the topographic reference from the actual faulty response. To perform a search by evaluation, the following stages in information transformation can be identified: from data

describing the failed state to the cause (what is changed in internal signals or functions) to the location of the faulty component. Thus, search by evaluation is a very similar process as heuristic diagnosis, in which not much search is carried out at all.

A very interesting feature of the study on electronic troubleshooters, as described in Rasmussen (1986) is the fact that technicians showed a pronounced ability to use general search routines that are not closely related to the specific instrument. Scanning a high number of observations by simple procedures is preferred to the preparation of specific procedures worked out by studying or memorizing the internal functioning of the system. Thus, these data indicate that people may have domain-independent diagnostic strategies that possibly can be used in a wide variety of domains.

According to Rasmussen (1986), diagnostic search strategies can be performed in basically two different ways. A set of observations representing the abnormal state of the system - a set of symptoms - can be used as a search template to find a matching set in a library of symptoms related to different abnormal system conditions. This kind of search is called *symptomatic search*. Symptomatic search consists of a form of pattern-matching between the symptoms and the result in terms of a label, which may be a cause, effect, location, or an appropriate control action. Symptomatic search is advantageous from the point of view of information economy, and a precise identification can frequently be obtained in a one-shot decision. Emphasis is put upon immediate knowledge, based on stored information. A serious limitation is that a reference pattern of the actual abnormal state of operation must be available, and multiple faults and disturbances may be difficult to take into account. Reference sets must be prepared by analysis or recorded from prior occurrences (role of experience!). Symptomatic search may be considered a form of heuristic reasoning as discussed before in the section about KADS.

The second type of search strategy is *topographic search*. The topographic search is performed by a good/bad mapping of the system through which the extent of the potentially "bad" field is gradually narrowed down until the location of the change is determined with sufficient resolution to allow selection of the appropriate action. The search depends on a map of the system that gives information on the location of potential sources of information for which reference information is available for judgments. The map is a model that identifies the potential sources of observations relative to the topology of the physical system itself. If the model contains information about functional aspects of the system, functional search may be carried out. Thus, functional search can be regarded as a form of topographic search. Information available in observations is used rather uneconomically by topographic search as described above, since emphasis is put on a search through deliberate knowledge. Topographic search may be considered a special type of systematic diagnosis as used by KADS, since in both systematic diagnosis and topographic search no direct relationship exists between symptoms and underlying faults.

All topographic strategies depend on search with reference to a *model of normal function* and are therefore well suited for identification of disturbances that are not empirically known. Symptomatic search strategies are in general very information-economic, but they can only be applied when a library of symptom to fault connections is at one's disposal.

Interesting in Rasmussen's approach is that the important distinction that he makes is between context-specific strategies (symptomatic search) vs. context-free strategies (topographic search). Thus, Rasmussen's distinction is based on the level of generality of search strategies. Context-free search strategies will be widely applicable, while symptomatic search strategies are tied to one specific context for their application. In AI-approaches it is more common to make a distinction between context-specific rules-of-thumb (heuristics) and, when heuristics are unavailable, the use of model-based reasoning, which is still domain-dependent reasoning, coupled with rather weak reasoning methods. Thus, AI-approaches are more geared towards a distinction between knowledge and search.

3.3.2 **Rouse**

The discrepancy between the supposed human abilities to react appropriately and flexibly in failure situations and the occurrence of "human error" by which failure situations may be aggravated has been one of the reasons for Rouse and coworkers to start a research program on human problem solving performance in fault diagnosis tasks. A series of experiments was carried out, and a number of mathematical models for human problem solving was developed, which are described in Rouse (1982), and Rouse and Hunt (1984). What these researchers were particularly interested in is the theoretical question whether diagnostic skills are context-free or context-specific, which has, among others, huge implications for training issues. Thus, their work shows similarities in this respect to Rasmussen's work.

The experimental work was done using four different fault diagnosis tasks. Two of them, TASK1 and TASK2, were context-free diagnosis tasks (computer simulations of graphically displayed network representations) in which no association exists between the tasks and a particular system or piece of equipment. These context-free network representation tasks enable diagnosis based on the structure of the network, and may therefore be considered the psychological pendant of model-based reasoning based on structural descriptions, as described in section 3.2.2. Since in these tasks used by Rouse there is no connection with a particular piece of equipment or system, it is impossible to develop context-dependent skills. However, one would like to know whether context-free skills can be used in context-specific tasks, and the question is whether subjects can be trained to have general skills that are transferable to context-specific situations. Therefore a third fault diagnosis task was devised, FAULT, in which hardcopy schematics of various systems, such as automobile, aircraft, and marine systems can be employed. In this way, context-specificity is ensured. Finally, a number of experiments was carried out using real equipment in which subjects were required to diagnose failures in four and six cylinder engines. Apart from the various tasks, different forms of aiding were constructed. The results of experiments carried out with these tasks showed that in general, context-free skills learned with computer aiding in one task can be successfully transferred to another context-free or context-specific task. Positive transfer of training can be explained as a reordering of priorities within a set of basic problem solving rules. This reordering appears to enable trainees to utilize the structure of the problem to a greater degree and thereby make more efficient tests in the sense of achieving greater reductions of uncertainty per unit cost. Performance on context-free tasks is highly correlated with performance on familiar real equipment tasks.

Based on the results of the experiments, Rouse and coworkers developed a number of models for human problem solving. The first model was the fuzzy set model, based on the idea that people may not always be able to determine with absolute certainty whether a certain component could cause a particular set of symptoms. The second model, the rule-based model, was based on the assumption that fault diagnosis involves the use of a set of heuristics from which a person selects, using some type of priority structure. However, the success of this rule-based model was limited to context-specific situations, in which subjects shift from topographic to symptomatic search, to use

Rasmussen's terminology. These two models were therefore integrated into the fuzzy rule-based model, which is a formalization of the dichotomy between symptomatic and topographic search. This fuzzy set rule-based model first attempts to find familiar patterns among the symptoms. If a match is found, symptomatic rules (S-rules) are used to map directly from symptoms to hypothesized failure. If there are no such familiar patterns, topographic rules (T-rules) are used to search the structure of the system. Thus, when heuristics are available, they are used. If not, subjects have to fall back onto a general search strategy such as topographic search, which is context-free. It should be noted that in Rouse's research, diagnosis is defined as the process from identifying symptoms to spotting the right cause for the malfunction.

Generally speaking, both the approach by Rasmussen and the approach by Rouse are geared towards identification of the different search strategies used by subjects in determining the fault given the set of symptoms. This may be considered a description of the *process* of diagnosis, an identification of procedures to carry out each diagnostic step. This approach may be complemented by an identification of the *task structure* of diagnosis, in which the different steps in a diagnostic process are identified, coupled with an ordering of those steps, in order to efficiently diagnose malfunctions. Such a specification of steps taken in a diagnostic strategy may also be viewed as a global diagnostic strategy. Such a model about the task structure of diagnosis in a technical domain will be described in the next section. In this model, not only the steps from identification of symptoms to the determination of the likely fault are discussed, but also steps towards compensation or correction of the fault. Thus, the model covers the task structure of troubleshooting as a whole: from identification of the system up until an evaluation whether the problem has been solved.

3.4 The task structure of diagnosis: a general model

In real-life diagnostic tasks one can distinguish between several discrete steps in the reasoning process. Based on a literature research and experience with the development of knowledge-based systems aimed at diagnosis, a model is proposed consisting of a number of steps. This model may be considered a specification of the task-level reasoning process (Wognum, 1990), a global strategy for diagnosis. The steps proposed show some similarity to the steps in diagnostic reasoning put forward by Wognum and

Mars (1989) and Wognum (1990), but are more extended, and further elaborated. The following steps can be distinguished:

- **Identification of symptoms**

  Symptoms are the first indication that there is something wrong. This does not necessarily have to lead to an alarm situation, but it may be just a process value that is moving towards the defined limits. "Being tired" is not necessarily a symptom for a certain disease, but when you think that you are now more tired than you were before, you should start wondering about possible causes. "The machine is slightly more noisy than yesterday". In this example "being noisy" is not the symptom, but "being more noisy" may possibly be. The reader may notice that one needs expertise to interpret a certain signal as a symptom. Symptoms can be hard to describe, especially when they have a large perceptual component. In the paper and board industry it happens quite often that operators debate about the appearance of *symptoms, or have very idiosyncratic names for symptoms,* which makes uniform communication about them rather difficult.

  Depending on the domain, symptoms may be easier or harder to detect. In industrial domains it is often quite obvious that there is a problem. On the other hand, for medical domains it may be a lot harder due to vagueness of presentation of symptoms.

- **Judgment: How serious is the problem**

  Depending on a judgment about the seriousness of the problem, the whole line of reasoning and action-taking may change. If a problem is serious, it is important to take some action right away, for example to save someone's life or to prevent that any other unsafe situations will appear. It may also include a prevention of halting of the installation. When the necessary actions have been taken, the normal diagnostic routine may still take over, but that depends on the situation.

  A correct judgment in this sense is very important in many domains, especially when human safety is at risk, but also in process control situations.

- **Determination of possible faults**

For a certain symptom or set of symptoms there is often more than one possible underlying fault. For quality problems in paper and board manufacturing, it is not unusual that there are more than thirty possible faults that may result in the symptom to diagnose. Especially when "tunnel vision" appears in situations of stress, it is likely that people "forget" to take some of those possible faults into account. Especially with shift work it occurs quite often that not all operators are aware of all possible faults, due to the fact that they have not been exposed to all of them.

- **Ordering of faults according to likelihood**

The ordering of possible faults according to likelihood is a process in which certain probabilities are "assigned" to each possible fault. In this way, an ordering of the list of possible faults appears, with the most likely candidates on top. It should be the input to the process of testing.

- **Testing**

Testing has the function of selecting the "right" fault out of many, by ruling out as many of the other candidates as possible. Testing often takes place by collecting additional evidence, and may, especially in humans, be more like "finding extra arguments for what you already thought it would be", than "find ways to rule out this possibility" ("confirmation bias").

- **Determination of repairs or remedies**

Repairs can take several forms. *Local* repairs apply to the exact fault at hand, and are meant to alleviate a specific problem. This is an example of *compensation* for faults (Rouse, 1982, cited in Bainbridge, 1984). *Global* repairs, on the other hand, seem to work in many situations, and can sometimes be applied without a full diagnosis of the problem. A local repair in the paper industry would be for example cleaning a dirty element. A more global repair would be slowing the speed of the paper machine down, since that seems to remedy many problems, but is in general not highly recommended. A global remedy in medicine may be "staying home for a day". This is an example of *compensation for symptoms* (Bainbridge, 1984).

- **Determination of the consequences of the application of repairs**

  In many domains it is important to realize what consequences the application of a repair might have. For example, what side-effects do certain medications have, or in a paper mill, does application of this repair imply that the installation has to be halted. The outcome of this reasoning process may have an influence on the choice of repair, if there is more than one possibility.

- **Evaluation: has the situation improved?**

  The final step in diagnosis would be an evaluation of actions undertaken in terms of improvement upon the situation. Has the problem been solved or are other actions necessary?

These steps constitute an outline of a general strategy for diagnosis. To complete each step, local strategies are needed especially suited for the completion of this particular step, together with underlying domain knowledge. The search strategies as defined by Rasmussen (1976) may be viewed as local strategies in this sense: a description of strategies used in fault finding.

It is expected that the ordering of the elements described above is, generally speaking, also the order in which the different steps of the diagnostic process are carried out. The model would therefore predict the following ordering:

Symptoms --> Judgment --> Possible faults --> Ordering --> Testing --> Determination of repair(s) --> Consequences of repairs --> Evaluation

However, in certain situations shunting is possible. Based on the judgment about the seriousness of the problem, it may occur that a global repair is applied right away, without any further diagnostic actions. This would lead to an ordering of

Symptoms --> Judgment --> Repair. This may of course be followed by a complete diagnosis, including the determination of faults and the application of local repairs. The model does not contain any reference to domain knowledge that is brought to bear in each step of the model, and is thus meant to be a domain-independent diagnostic strategy. To solve a specific diagnostic problem also implies that the categories should be filled in by using specific knowledge about the domain. However, depending on the knowledge a subject brings to bear on a specific diagnostic situation, it may be that model-based reasoning has to be used to infer the specific content of each category, and it is therefore quite likely that the protocols will show pieces of model-based reasoning at any stage of the reasoning process whenever heuristic knowledge is insufficient.

The next deviation of the model may be the moment that the judgment appears in the protocol. According to the model it should appear early on in the diagnostic process, but the question is, if this judgment is verbalized later on, whether the subject only realizes it at that point that such a judgment should be made, or whether he has made this judgment early on, but verbalizes it later. Therefore, judgment may appear anywhere in the protocols.

Since it happens quite often that symptoms can be caused by many possible faults, it is likely that at least some testing occurs before the enumeration comes to an end, to avoid working memory overload. This implies that transitions Possible Fault --> Testing are legitimate ones.

In the paper and board industry, the determination of repairs is often a rather trivial matter, once the fault has been determined, since it often implies cleaning or replacing the faulty part. Therefore, it would not be surprising if repairs are not mentioned. In this respect, the transition Testing --> Consequences is a legitimate one.

Finally, due to limitations of working memory, it is always possible that the subject goes back to the previous category, since it may be that he is not able to keep all the information active in his working memory. Thus, the model should allow backing up to the category right before the present category.

The ordering of categories in the model may be specified by summing up the *legitimate transitions* (Hamel, 1990). For each category, the categories that are permissible to follow it are specified. This results in the following matrix, depicted in table 3.6.

Table 3.6 *Legitimate transitions in the diagnostic model.*

from				to				
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*	*			*		
Judgment	*		*	*	*	*	*	*
Faults	*	*		*	*	*		
Ordering		*	*		*	*		
Testing		*	*	*		*	*	
Repairs		*			*		*	
Conseq.		*				*		*
Eval.		*					*	

3.5 Conclusions

Unlike the KADS diagnostic model, the model presented in section 3.4 is not a competence but a performance model of diagnosis, since it contains assumptions about underlying knowledge and limitations of working memory. Another difference with the KADS interpretation models is that the model presented here is more elaborate and less sketchy, and therefore a stronger model for diagnosis which will be easier to refute. Unlike the models based on structure and behavior of devices, it is a psychological model, aimed at describing human performance. Also, the model presented here and the structural models have a difference in emphasis: the model presented here is a task-level description of diagnosis, while the model-based reasoning approaches are more focussed upon a specification of underlying knowledge.

The main difference with the psychological models presented is the fact that the model presented here contains a task-level description of diagnosis, while the psychological models by Rasmussen and Rouse aim at a description of local strategies at the domain level (Wognum, 1990): strategies used to complete the steps specified at the task level. Thus, in comparison, the two models are aimed at different levels of diagnosis: the task level versus the domain level. Both the models by Rasmussen and Rouse only describe local strategies at one part of the task level: fault finding. To obtain a more complete idea about local strategies used at other parts of the task level, more research is needed. However, this research topic falls outside the scope of this study.

In conclusion, a task-level model for diagnosis is presented, which should suffice to describe the whole process from the identification of symptoms to the undertaking of compensatory actions. This model may also be viewed as a description of the global steps taken in diagnosis: a global strategy.

It is expected that experts perform more or less according to this strategy. The model will serve as a scoring and interpretation scheme for the protocols that have been obtained from experts and novices solving alarm situations in a technical setting. In this way, evidence can be obtained about the correctness and completeness of the model, and the usefulness of the model as a psychological model of human behavior. As stated before, expert operators are supposed to behave as the model prescribes, and thus, in this situation, it is expected that the expert operators follow the diagnostic strategy as outlined above. With respect to novices' behavior the question is whether the model will give an accurate account of their diagnostic strategy, or whether qualitative differences between experts and novices can be found, that are not solely attributable to a mere lack of knowledge that hinders the performance of novices. Thus, the following questions will be posed: does the proposed model give a good account of both expert and novice behavior in diagnostic situations, and, if not, what elements in the model should be modified such that it gives a good account of the observed behavior. Another question is whether evidence can be obtained for the existence of a general diagnostic strategy, which is not completely dependent upon the availability of underlying system knowledge. The following chapters will discuss experiments relevant to these questions.

Chapter 4

Content and structuring of the declarative knowledge base

4.1 Introduction

In a large number of domains such as chess (Chase & Simon, 1973a, 1973b; De Groot, 1946), Go (Reitman, 1976), baseball (Chiesi, Spilich, & Voss, 1979), bridge (Engle & Bukstel, 1987; Charness, 1979), music (Sloboda, 1976), soccer (Morris, Gruneberg, Sykes, & Merrick, 1981), electronics (Egan & Schwartz, 1979), physics (Larkin, McDermott, Simon, & Simon, 1980), and computer programming (McKeithen, Reitman, Rueter, & Hirtle, 1981; Barfield, 1986) differences are found between experts and novices in their abilities to recall large amounts of briefly presented meaningful information. A common explanation for this difference is that experts not only have *more* information, but that they also have that information better organized into useful bigger chunks. Instead of perceiving and remembering individual pieces of information, they process meaningful groups of information, making their perception more efficient and their recall performance much higher. Superior organization appears to be the key to expert performance, coupled with strong domain-dependent strategies.

However, one of the questions that still remains concerns the interaction between knowledge one can bring to bear in a specific task and the development and utilization of specific task-dependent skills (Greeno & Simon, 1988). Relevant to this question is the recent study by Boshuizen (1989), who found that with increasing expertise, physicians develop "illness scripts". In these illness scripts the conditions of the patient are described, together with his or her medical, hereditary or social background that may have contributed to the present disease (Enabling Conditions), the disease process itself (Fault), and the signs and symptoms that are caused by the disease and the course it may take (Consequences) (Feltovich & Barrows, 1984). In fact, these illness scripts may be taken as evidence for a well-structured declarative knowledge base, that can be used in diagnostic problem solving situations. This should be coupled with efficient problem solving strategies, geared towards diagnosis. The model presented in chapter Three may be regarded such a strategy. Boshuizen (1989) showed that these illness scripts do not originate solely from knowledge compilation and enrichment since the younger students in her study already applied rudimentary illness scripts when diagnosing a case, which implies that the development of strategies and the organization of knowledge are somehow independent processes.

In the study presented here one of the questions that will be addressed concerns the interaction between the amount and organization of domain-specific declarative knowledge and the development of diagnostic skills in the domain of paper making. In the present chapter results will be discussed of an experiment carried out to gain insight in expert-novice differences with regard to the content and organizations of the declarative knowledge base. Chapter Five and Six will be geared to the strategies employed in diagnostic tasks, based on a comparison between thinking-aloud protocols of persons of different levels of expertise and the generic diagnostic model developed in chapter Three. Chapter Seven will again continue on the relationship between domain-specific system knowledge and problem solving strategies. Finally, chapter Eight presents an overall discussion of the findings related to diagnosis.

To gain insight into the content and structure of the declarative knowledge base of expert and novice paper makers an experiment was carried out to investigate whether there are differences between experts and novices with regard to one important aspect of diagnosis; the generation of possible faults for a class of symptoms: paper breaks. The subdomain of paper breaks was chosen firstly because paper breaks occur rather frequently in a paper mill, so all subjects would have at least some experience in diagnosing them. Second, paper breaks are known to occur for many different reasons. Therefore, generating possible faults should not be too difficult.

4.2 Method

4.2.1 Subjects

Twenty subjects participated in this experiment, including nine experts with more than five years of experience as independent process operator (mean number of years of experience = 15), eight trainees with less than three years of experience as operator, and not yet able to run the installation independently, and three novices who were employed by the paper mill for two years. They have only received a general training on all aspects of paper making. All of the subjects have participated in other experiments reported in other chapters, depending on their field of expertise in either the scenario-experiment or the simulator-experiment, and some of the experiments that will be reported on the availability of system knowledge.

4.2.2 Procedure

Subjects were asked individually, in a quiet setting, to mention as many possible faults for paper breaks as they could. They did not know beforehand that this question would be asked to them, so no preparation could have taken place. The answers were recorded on tape, for which the subject had given his permission in advance of the whole set of experiments. The subjects were told that the results were to be regarded as confidential: no information about the performance of a specific subject would be revealed to anyone else.

Subjects were given as much time as they needed to complete their enumeration. To prevent discussion among subjects afterwards, all expert operators and trainees were run as quickly as possible, within one week (shiftwork). The novices were run at a later time. However, this is not considered to be a problem since neither of them had much contact with any of the other subjects at all.

4.3 Results and discussion

4.3.1 Content of the knowledge base

Table 4.1 gives an overview of the mean number of possible faults for paper breaks mentioned by experts, trainees, and novices.

Table 4.1 *Mean number of possible faults for paper breaks mentioned by experts, trainees, and novices.*

Experts	Trainees	Novices
35.2 (n = 9)	34.5 (n = 8)	23.7 (n = 3)

According to the Kruskal-Wallis Test Statistic T, there is no difference overall between the three groups, T = 1.63, n.s.

This result implies that no differences exist between the number of possible faults mentioned for paper breaks between the three groups. However, the absence of a significant difference between the novices and the other two groups may partly be due to the small sample size of novices. Thus, although it is rather likely that there is in fact a difference between experts, trainees, and novices with regard to number of faults mentioned, this may not show up due to the small number of subjects. On the other hand, although novices perform at a lower level than the two other groups (30% lower), they are still able to come up with a reasonably large number of possible faults, and in fact, they perform at a very reasonable level. Chapter Five and Six will discuss this finding in more detail.

4.3.2 Organization of the knowledge base

The order in which subjects produce possible faults reveals information about the manner in which information is stored in memory and about the knowledge structure that governs the storage and retrieval process. To obtain this ordering, all items in each protocol were, if possible, assigned to a category. Examples of these categories are "faults in the wire section", "faults in the press section", "mechanical defects", and "fouling". Since the categories that occur in protocols could not be known beforehand, the result of this ordering is an assignment of each item in the protocol to either a category to which more items in the protocol belong, or an assignment of a particular item as belonging to a category "single". To ensure some objectivity, the assignment of items in the protocols to categories was carried out by two persons, who were both knowledgeable about the subject of paper breaks. The level of agreement about the assignment of items to categories was high, exceeding 90%.

A measure for the amount of organization was found in the Adjusted Ratio of Clustering (*ARC*) (Roenker, Thompson, & Brown, 1971). This measure has also been used by Claessen and Boshuizen (1985) and Boshuizen (1989) and is generally regarded as a satisfactory measure of category clustering in free recall (Murphy & Puff, 1979). *ARC* scores express the degree to which items of the same category are grouped together in a recall protocol, as opposed to randomly dispersed throughout the listing. Chance clustering is set at zero, and perfect clustering at one. Thus, the *ARC* score represents the proportion of actual category repetitions above chance to the total number of possible category repetitions above chance for any given recall protocol. Hence, the *ARC* score is invariant with respect to factors unrelated to relative amount of clustering. The computational formula for the *ARC* score is as follows:

$$ARC = \frac{R - E(R)}{Max(R) - E(R)}$$

where

R = total number of observed category repetitions (i.e. the number of times a category follows an item from the same category),

$Max(R)$ = maximum possible number of category repetitions, and

$E(R)$ = expected (chance) number of category repetitions.

It should be noted that

$$Max(R) = N - k$$

where N = total number of items recalled, and k = number of categories represented in the recall protocol. And

$$E(R) = \frac{\sum_i n_i^2}{N} - 1$$

where n_i = number of items recalled from Category i, and N is as before. To obtain *ARC* scores for each subject the "single" items were not treated as categories. The reason for this is that without predetermined categories it is not clear whether "single" items constitute tokens of a category or a whole category represented by one exemplar only.

Table 4.2 gives an overview of the mean *ARC* score for the three groups of subjects.

Table 4.2 *Mean ARC scores for experts, trainees, and novices.*

Experts	Trainees	Novices
0.61 (n = 9)	0.53 (n = 8)	0.44 (n = 3)

According to the Kruskal-Wallis Test Statistic T, there is no difference overall between the three groups, T = 1.91, n.s. Since the variation in both *ARC* scores and number of items recalled is quite high between subjects, Pearson R was computed to investigate whether there is a relation between number of items recalled and *ARC* score. This yields a significant result: Pearson R = 0.58, $p < 0.01$, implying that the higher the *ARC* score, the more items recalled. In other words, a better organization of knowledge goes together with better recall. Since it is expected that a better organized memory leads to fewer isolated items, a correlation was computed between percentage of "single" items in the recall protocols and *ARC* scores (again, no differences between groups could be found), yielding a significant result, Pearson R = - 0.58, $p < 0.01$.[1]

To summarize these results: no differences were found in the number of items recalled between groups with different levels of expertise, although these results may

---

[1] The fact that both the correlation between number of items recalled and *ARC* score, and the correlation of percentage of "single" items and *ARC* score turn out to be .58 (.58 and -.58) should be taken as coincidence.

partly be due to the small sample size of the group of beginners. Also, no difference was found between degree of organization as measured by the *ARC* score and level of expertise. However, a relationship exists both between percentage of "single" items and degree of organization, and between degree of organization and number of items recalled. This result implies that level of expertise is *not* the critical factor in number of items recalled, but rather degree of organization of knowledge: the better memory is organized, the more one recalls. This implies that there are significant individual differences within one level of expertise in the way memory is organized, or in recall strategies. The question remains however, how well this knowledge can be brought to bear in diagnostic problem solving situations. Experiments concerning this topic will be reported in chapter Five and Six.

Chapter 5

Strategies for diagnosis: the scenario experiment

5.1 Introduction

Many researchers make a distinction between declarative and procedural knowledge (e.g. Anderson, 1983). Declarative knowledge may be conceived as a collection of stored facts and is also called system or device knowledge in the domain of technical systems. Procedural knowledge can be regarded as a collection of actions or procedures that an intelligent system can carry out. It also contains knowledge of the procedures with which one investigates a device to make diagnoses about its functioning. Procedural knowledge is content-specific and thus only applicable in a limited domain. In addition to the declarative-procedural distinction, a distinction can be made between domain-specific knowledge and strategic or metacognitive knowledge. This strategic or metacognitive knowledge is knowledge about general principles of thinking and reasoning that is applicable across specific content domains. With this decomposition, it is assumed that *procedural* and *device* knowledge are organized and deployed by goals, plans and decision rules that comprise *strategic* knowledge. Thus, strategic knowledge can be said to have a *control* function to enable dynamic, flexible reasoning. As described in Gott (1989), support for the concept of strategic control knowledge comes from a number of academic domains such as geometry (Greeno, 1978), text editing (Card, Moran, & Newell, 1983), computer programming (Anderson, Boyle, Farrell, & Reiser, 1984), simple device operation (Kieras & Bovair, 1984) and electronic troubleshooting (Gott & Pokorny, 1987). The question is though, whether this strategic knowledge, being general in nature, can be transferred to tasks in related fields, in this manner enabling intelligent systems to generalize over domains. An example of a recent attempt to build a system that is able to generalize over domains, and in which the interaction between domain specific and domain independent strategic knowledge is explored, is Fermi, an expert system that reasons about natural science domains (Larkin, Reif, Carbonell, & Gugliotta, 1988). In Fermi, domain specific knowledge of scientific principles and strategic problem solving knowledge are encoded in separate but related semantic hierarchies. This allows the system to apply common problem solving principles of invariance and decomposition as encoded in the strategic problem solving knowledge base to a variety of problem domains such as fluid statics, DC-circuits, and centroid location. Similarly, in the knowledge based tutoring system Guidon for a medical domain (Clancey, 1987), it turned out to be very important to separate tutoring knowledge, and knowledge about general strategies, from specific domain knowledge. Processes of acquiring strategic knowledge have been addressed in analyses by Anzai and Simon (1979) on the Tower of Hanoi and by Anderson, Farrell, and Sauers (1984) on the acquisition of knowledge needed to learn Lisp. As stated in Greeno and Simon (1988), both studies showed that important factors in acquiring strategic knowledge are the activation of a problem goal that can be achieved by a sequence of actions and the acquisition of productions in which the action of setting the goal is associated with appropriate conditions in the problem situation. The importance of strategic knowledge has been shown by Greeno (1978). He conducted an experiment on the acquisition of high school geometry knowledge. As the computational model Perdix of problem solving behavior showed, strategic knowledge is needed for setting goals that organize problem-solving activity. One of the students in Greeno's study knew the problem-solving operators and the geometric patterns to achieve them, but was unable to solve the problem. This result can be explained by the hypothesis that the student lacked knowledge about the problem solving strategy needed in this problem. Schoenfeld (1979) showed that students who were given explicit instruction in the use of heuristic strategies in the domain of college mathematics, showed superior performance compared to students who had not received this training. As the reader may have noted, a difference exists between strategic knowledge as implemented in Perdix or as trained in Schoenfeld's study, and strategies as used by, for example, GPS (Ernst & Newell, 1969) or Soar (Laird, Newell, & Rosenbloom, 1987). In GPS and Soar, only very general problem-solving strategies, or so-called weak methods such as means-ends analysis, have been implemented. These weak methods do not contain any domain-specificity. In Perdix and in Schoenfeld's study the emphasis is put upon a generic strategy for a class of problems, such as diagnosis or geometry problems, which is a rather different enterprise.

The question remains, though, how this strategic knowledge is acquired, and what the interaction is between strategic knowledge and device knowledge. In technical domains such as paper making, an important source of strategic knowledge may be a strategy for solving diagnostic problems, such as the one described in chapter Three. Can evidence be found for a general diagnostic strategy used by expert operators and what are the differences between experts and novices with respect to the diagnostic strategy applied? The following experiment was designed to shed light on this issue. A comparison will be made between troubleshooting strategies of expert operators, novices, and a transfer subject: an expert paper maker confronted with novel problems (different domain). It is also investigated what types of system knowledge are brought to bear when handling diagnostic problem-solving situations. The task-level diagnostic strategy as developed in chapter Three will serve as a guideline for interpreting the data. To be able to obtain the best possible insight into the subjects' mental processes, the technique of thinking-aloud protocols will be used. These protocols provide a much more detailed description of the cognitive processes than just the final answer that a subject gives. The use of protocols enables inferences about intermediate steps such as subgoals, and attention to specific aspects of the problem. Protocol statements are not treated as introspective descriptions of psychological processes, but rather as overt reports of mental activity that the subject is aware of, but usually does not verbalize. For a complete overview of the technique of protocol analysis, the reader is referred to Ericsson and Simon (1984).

The analyzed protocols will be compared to the general diagnostic strategy as proposed in chapter Three in order to obtain a psychological model for diagnostic behavior.

First of all, an analysis will be given of the goodness of fit of this strategy as compared to expert performance. This analysis will try to answer two questions: are the diagnostic categories from the model also used by experts, and secondly: are experts and novices alike in meeting the proposed ordering of categories from the model? This will be followed by an in-depth analysis of the behavior of the three groups of subjects. Special attention will be given to the transfer subject: does he, since he is posed with novel problems, have to fall back on general problem solving strategies such as means-ends analysis (does he show novice behavior), or is he able to use a more specific diagnostic strategy, even though he lacks domain knowledge. If he is able to use a specific diagnostic strategy in these novel situations, than this may be taken as evidence for transfer possibilities of this strategy to at least different subdomains of paper making.

5.2 Method

5.2.1 Subjects

Five subjects participated in the experiment, including one expert (five years of experience in stock preparation), three novices who are the same as in the experiment on paper breaks reported in chapter Four, and one transfer subject. The transfer subject is an expert paper maker with more than ten years of experience as independent operator. However, in the area of stock preparation his experience is very limited. These subjects will also participate in some of the experiments reported in chapter Seven.

5.2.2 Materials

Based on an analysis of possible alarms occurring in stock preparation the following alarm situations were selected:

1. Breakdown of conveyer belt of pulper 21 or pulper 1
2. Low consistency of stock in chest K2a or K21
3. Clogged cyclone
4. Clogged fiberizer 2F2 or 1F2
5. High level indication of stock chest 23 or stock chest 12
6. Failure of pump of stock chest 25 or stock chest 14

All selected alarm situations are frequently occurring alarms and were judged by an independent shift manager as being not too difficult to solve and representing a wide range of complexity of diagnosis for an operator. For example, if problem six occurs there is not much an operator can do in terms of keeping the stock preparation going, apart from calling a technician to repair the pump. However, problem two or problem four require a lot more activity from the operator, since it is certainly possible under certain conditions to keep producing while the fault is being repaired.

5.2.3 Procedure

All subjects were run individually in a quiet location not close to the actual stock preparation department. They were asked to think aloud while diagnosing the alarm situations (scenarios). Subjects were given the following instruction: "Imagine you're the operator responsible for stock preparation. The following alarms will occur. Please tell me as complete and as exactly as possible what you would do in this situation." Some practice in thinking aloud and the type of question posed was given before the experiment started. The whole session was recorded on tape, for which the subject had given his permission in advance of the experiment. The subjects were told that the results were regarded as confidential: no information about the performance of a specific subject would be revealed to anyone else.

All subjects received all scenarios in the same order, and they were told by the experimenter that they could use as much time as they needed to solve the problems. All of them finished within thirty minutes.

5.3 Results and discussion

All protocols were transcribed verbatim and, based on pauses in speech, divided into statements. The protocols were scored according to the categories of the general diagnostic model as put forward in chapter Three. One additional category was added, "comments", which was used for task-irrelevant comments by the subject. The scoring was done independently by two persons, one being the experimenter, and the other one a psychologist with sufficient knowledge about paper making, and therefore acquainted with the concepts used by the subjects. She was not aware beforehand whether a certain protocol belonged to a novice, the expert, or the transfer subject. Even without discussion there was a high level of agreement (85%) upon the assignment of categories to statements in the protocols.

It turned out that 92% of the statements could be assigned to categories of the diagnostic model, which gives confidence in the completeness of the model.

Table 5.1 gives an overview of the important results from this experiment as extracted from the protocols summarized over all six alarm situations. Special attention is given to the most important categories from the diagnostic model. The results of the three novices have been averaged over subjects.

Table 5.1 *Categories from the diagnostic model, summarized over all six alarm situations, for all three groups of subjects. The results from the novices have been averaged over subjects.*

	Expert	Transfer	Novices
Number of faults	15	12	4.5
Number of repairs	15	11	15.3
Judgment (max = 6)	4	6	1.3
Number of categories filled in, summarized over all 6 problems (max = 48)	32	29	15
Model-based reasoning	5	10	9

In chapter Four, it was concluded that both experts and novices are able to perform at a reasonable level when asked to enumerate possible faults for a class of symptoms: the novices came up with 30% fewer faults than the other two groups. However, in that experiment, the selection of possible faults was not embedded in a "real" alarm handling situation. The results presented here show a different picture: there is a rather big difference between the expert and transfer subject on the one hand, and the three novices on the other hand, with the novices performing poorly compared to the other subjects. In this experiment a difference of 70% between experts and novices is found, compared to 30% in the previous one! The Mann-Whitney U Test Statistic revealed a marginally significant difference with respect to number of faults mentioned between the expert and transfer subject on the one hand and the novices on the other hand, U = 6, p = .08. The fact that these results are only marginally significant is presumably due to the small sample size. This result suggests that declarative knowledge about paper making, although available, is not activated in the alarm handling situations. This finding is in line with results obtained by Feltovich (1981), who found that in a study on expert and novice performance in medical diagnosis, that novices mentioned significantly fewer hypotheses during problem solving than experts.

Another interesting result is the number of repairs mentioned by the three groups. In this case, the novices perform at the same level as the expert and transfer subject. However, there exists an unbalanced situation between the number of faults mentioned and the number of repairs, which is worth investigating. This aspect of the novices' protocols will be discussed more thoroughly later on in this chapter.

If the overall quality of the diagnostic reasoning processes of the three groups is taken into account, as measured by the number of categories from the diagnostic model filled in, again there is a big difference between the expert and transfer subject compared to the three novices, Chi-square = 106.0, df = 4, $p < .001$. It should be mentioned, that the expert and transfer subject perform better than may appear at first sight. One category from the diagnostic model, Evaluation of improvement, will not appear in the protocols obtained in this situation, since an evaluation of improvement can only be given in a situation in which one is really in a position to perform the proposed repair, and this is not the case in the hypothetical situations presented. Therefore, the results are deflated. A remarkable result with respect to the number of categories that is used is that at least one of the categories, Judgment of the seriousness of the alarm, is almost completely lacking in novices' protocols. In one occasion, this judgment is made only on the last problem, in which it is almost unavoidable to make this judgment, since a pump failure in chest 25 or chest 14 causes the whole production line to go down. Chi-square on Judgment revealed a significant difference between the expert and transfer subject on the one hand, and the novices on the other hand, Chi-square = 14, df = 3, $p < .01$. This is a rather important omission in novices' protocols, since, as explained in chapter Three, the outcome of this judgment determines what needs to be done next: a serious problem asks for an immediate (global) repair.

The transfer subject appears to behave more like an expert than a novice, which is somewhat surprising since he, just as the novices, does not have much practical experience in stock preparation. The data show that both the transfer subject and the three novices have to use a substantial amount of model-based reasoning, such as reasoning based on process flow, to give their answers. This is not surprising since they are not expected to have very detailed system knowledge in this area. A more detailed discussion below of the strategy followed by the transfer subject, will show however, that his reasoning strategy is rather similar to the experts' diagnostic strategy. Thus, it appears that he does not have to fall back on weak methods for his problem solving strategy, but is able to rely on more powerful general heuristics for diagnosis. On the basis of these data we may tentatively conclude that transfer of the diagnostic strategy to the novel situation has occurred.

Apart from the categories from the diagnostic model used by the subjects as such, it is also important to look at the order in which those categories are used, since that provides another test of the model. In chapter Three, it was specified what are legitimate transitions in the model. In what follows an overview will be presented of the transitions in the protocols of the expert operator, the transfer subject, and the novices.

Table 5.2 *Transitions in the protocols of the expert subject. Legitimate transitions are printed in italics.*

from	to Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*1*	*2*			*4*		
Judgment						*1*		
Faults				*1*	*4*	*5*		
Ordering						*1*		
Testing			*1*			*3*		
Repairs		*3*	4		*2*		*5*	
Conseq.			1			*2*		
Eval.								

The total number of transitions in the expert's protocol was counted to be 40, of which only five were illegitimate (12.5%), falling into two types: Repair --> Fault, and Consequence --> Fault. It is interesting to investigate what happens in the protocols when these deviations of the model occur. From looking at the protocols it turned out that what the expert in fact does is reasoning in a depth-first way: he fully completes his line of argumentation for a certain fault, including the repairs and the consequences, before taking up another possible fault. This may explain the illegitimate transitions: at the moment his reasoning about a certain fault comes to an end, he starts with the next possible fault. The following excerpt from his protocol on problem five illustrates this point:

-	*It may be caused by a wrongly adjusted level* ** of chest 23 **..	** Fault **
-	*Usually, if the level has been adjusted wrongly, you have to correct it..*	** Repair **
-	*And if it is high for any other reason* ** chest level **	** Fault **
-	*Then you have to stop the pump of chest 22A..*	** Repair **

What is interesting to note in the protocols of the expert subject is that, although he shows quite a few transitions Symptom --> Repair (four times), the problems on which these transitions occur are always judged by the subject to be serious problems in which quick actions have to be undertaken .

Table 5.3 shows the transitions in the protocols of the transfer subject.

Table 5.3 *Transitions in the protocols of the transfer subject. Legitimate transitions are printed in italics.*

from	to							
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*2*	*1*			*2*		
Judgment			*1*		*1*	*3*		
Faults		*1*			*1*	*1*		
Ordering								
Testing		*1*	*1*			*1*		
Repairs		*5*					*2*	
Conseq.			1					
Eval.								

The total number of transitions is 24 of which 23 are legitimate transitions. The only deviation is one instance of Consequence --> Fault, which may be explained in the same way as for the expert: the transfer subject uses depth-first reasoning. Similar to the expert, whenever Symptom --> Repair transitions occur in the protocols of the transfer subject, they are always accompanied by a judgment that the problem is serious.

Finally, in table 5.4 the results of the novices will be presented.

Table 5.4 *Transitions in the protocols of novices. Legitimate transitions are printed in italics.*

from	to							
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*1*	*3*			*13*		
Judgment						*1*		
Faults					*1*	*6*		
Ordering								
Testing			*1*			*2*		
Repairs		*3*	7		*2*		*2*	
Conseq.			1			*1*		
Eval.								

Out of the total of 44 transitions, 8 can be considered illegitimate. They can all be traced to belong to the same sort as those discussed before from the expert and transfer subject: depth-first reasoning. Noteworthy in the behavior of the novices is that Symptom --> Repair transitions are *not* accompanied by a judgment about the seriousness of the problem, which is an important difference with the transfer and expert subject. The examples taken from protocols from experts and novices on pages 83 and 84 will clarify this point.

Summarizing these results, it is fair to state that the empirical results do not contradict the model proposed in chapter Three. The only deviation found is the use of depth-first reasoning by all subjects, which was not explicitly predicted by the model. One should also note that the category Evaluation is never used by any of the subject, presumably since the scenarios remain verbal problems and one can only evaluate after some action has been undertaken. Another category that does not show up very often is an explicit ordering of the possible faults in terms of likelihood. A possible explanation for this may be that this is a very implicit process and therefore is not likely to show up in the protocols.

In what follows, a more qualitative analysis of the results of the expert, the three novices, and the transfer subject will be presented and subsequently discussed. Special attention will be given to the aspects of the diagnostic model mentioned above.

Diagnostic behavior of an expert in the area of stock preparation

One of the questions posed in the introduction of this chapter was what kind of strategy an expert is using while diagnosing alarm situations. Does he use a diagnostic strategy as put forward in chapter Three or does he follow a completely different line of reasoning? The model for diagnosis put forward in chapter Three was suggested to have normative aspects, in the sense that the more categories of the model would be filled in in a certain diagnostic situation, the more systematic, and therefore, better, it would be. Second, the model contains a logical ordering for setting up goals and subgoals. The first point (how many categories of the model have been filled in) was answered before: in six problems, experts filled in 32 of the 48 (six times eight) possible slots of the model. The second question, how well is the logical ordering obeyed, was answered above by presenting the transition tables for the different groups of subjects. It will also be illustrated by the following coded and somewhat stylized prototypical protocol of the expert on problem 2. Statements of the experimenter, as well as the assigned categories are preceded by **. Before the protocol excerpt will be presented, figure 5.1 is included to obtain insight in the process flow around buffer chest 21 (K21).

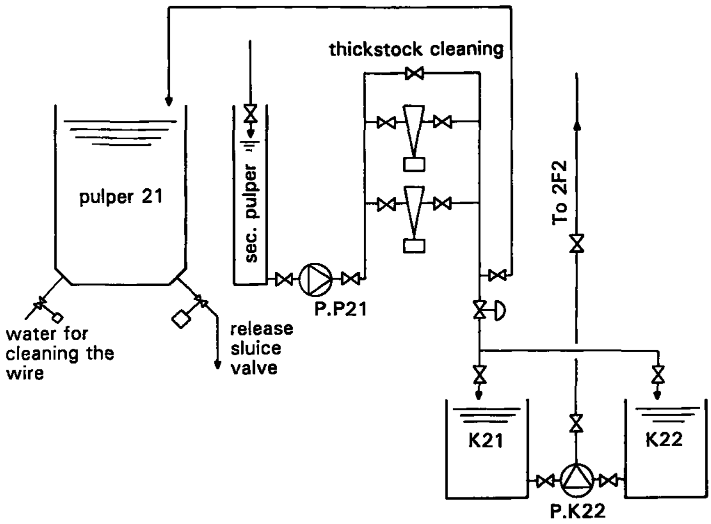


Fig. 5.1 The process flow around buffer chest 21 (K21).

Protocol of expert on problem 2: Low consistency of stock in chest 21

**	*Allright, let's take the second alarm.. low consistency of stock in chest 21..*	
-	*stock in 21 is too thin..*	** symptom **
-	*first of all, see whether stock in the pulper has the right concentration..*	** possible fault ** ** testing **
-	*as you know.. stock goes from the pulper to chest 21*	** model-based ** explanation
-	*is that all right.. and the secondary pulper is not empty*	** possible fault **
-	*this means.. if the secondary pulper is empty there is a problem with the the pump since only water is*	** possible fault **

	pumped..	
-	*this means that the wire in the secondary pulper has stopped-up..*	** possible fault**
-	*which is often caused by broke paper*	** possible fault **
-	*but it might also be caused by water on the release sluice valve*	** possible fault **
-	*and since recently, it is possible to use water for cleaning the wire.. this valve may wrongly be opened*	** possible fault **
-	*but that you will realize directly because of the pulper level*	** testing **
-	*if this is not the case.. then clean the pulper..*	** repair **
-	*since it has become too filthy..*	** possible fault **
-	*then the problem will surely be alleviated.. no doubt.. this is not a very serious problem..*	** judgment **

The following abstracted diagnostic reasoning scheme may make even clearer what line of reasoning the expert actually follows:

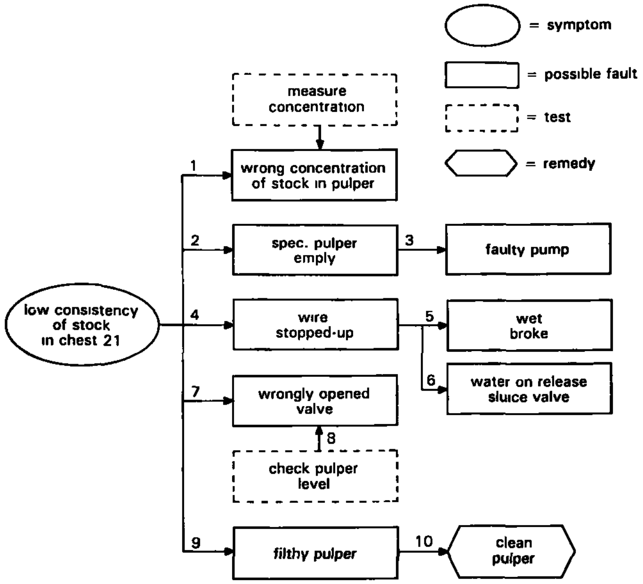


Fig. 5.2 Abstracted diagnostic reasoning scheme of expert reasoning.

Five possible faults are mentioned:

1. Wrong concentration of stock in pulper (inferred from first statement)
2. Secondary pulper empty, which implies a faulty pump
3. Wire of secondary pulper stopped-up, which may be caused by wet broke or water on the release sluice valve
4. Wrongly opened valve
5. Filthy pulper

For fault two and three the expert mentions a chain of faults. This is rather characteristic for his line of reasoning which is aimed at finding the exact solution to the problem and only applying a global repair when it is really necessary.

In addition, it should be noted that, based on statement two and three, it is inferred that the subject made an ordering of possible faults for this symptom. Although not explicitly mentioned by the expert, each specific fault implies a repair. One of the reasons for the expert not mentioning these repairs is that they are rather obvious to him. This appears to happen quite frequently in the paper and board industry. From experience with knowledge elicitation for expert systems on diagnostic problems in this type of industry it is known that finding the right fault for the symptom is the hardest part of the diagnostic process. However, when the right fault has been spotted, finding the right repair is a rather trivial matter, since it often implies either cleaning or replacing a certain element (Schaafstal & Vos, 1988). Implicitly, there is also an ordering of possible repairs in the expert's protocol: the pulper only has to be cleaned when all the other possible faults have been excluded. The example of model-based reasoning does not play a role in the reasoning process itself, it only serves the function of explaining something to the experimenter, but nothing is deduced based on this line of reasoning.

As the judgment implies: this is not a very serious problem, and therefore there is quite some time available to diagnose the problem, and apparently, it is judged to be worth the time and effort. This does not apply to all types of alarms. For example: the protocol on problem 3 shows a rather different situation:

Protocol of expert on problem 3: Clogged cyclone

-	*OK.. clogged cyclone..*	
-	*replace this one*	** repair **
-	*immediately..* *there is always one on reserve..*	** judgment **
-	*then clean it..*	** repair **
-	*if there is time and you have* *the possibility then open the* *cyclone and clean it.. with* *two people..*	** repair **
-	*If you cannot get it opened..* *or you don't have the time..* *then call the maintenance staff*	** repair **

'cause otherwise it will take too long.. ** judgment **

This protocol shows that in certain cases the expert is not necessarily looking for the fault, but immediately has to react to the situation, based on his judgment how serious the alarm is. It also shows another interesting aspect of his behavior: how to repair is carefully planned: if the facilities to repair are not readily available (no time or no manpower) someone else is called who will presumably be better equipped. This point is related to the division of responsibilities between the different groups of employees: the person responsible for stock preparation is supposed to do some diagnosis and repair of alarms, but only when the repair does not take too much time. He always has to keep in mind that his prime responsibility is stock preparation and not technical repair or maintenance, and thus, he has to set his priorities.

Statements that can be taken as evidence for model-based reasoning from the expert are of the following types:

1. Reasoning based on the process flow
   For example in problem four:

   .. *I would stop the pump of chest 22 A.. then I have to ensure nothing gets into chest 22A.. nothing in chest 22.. so I should not supply anything to the pulper.. and all this as an effect of stopping the pump of chest 22A..*

2. Reasoning about the topographical location of parts of the process
   For example in problem six:

   .. *Pump 25.. there is not much I can do.. is located near the cabin of the paper maker.. I only get the announcement..*

3. Reasoning about how the process is controlled
   For example in problem five:

.. If there is something wrong with the automatic valve between chest 23 and 24 then the level of 23 will rise..

Out of the total of five pieces of evidence for model-based reasoning, there is one of type 1, two of type 2, and two of type 3. No evidence could be found for deep reasoning based on the function of the installation or reasoning with regard to the influence of the faulty piece on the quality of the paper or pulp.

Novices' diagnostic performance in the domain of stock preparation

From table 5.1 it was concluded that novice operators almost never make a judgment about the seriousness of the problem, although in fact it is a rather important judgment to make. In expert behavior it changes the diagnostic strategy, especially the order in which goals are set up. The following excerpt from a novice protocol shows what kind of behavior occurs when someone does not make this judgment:

Novice protocol on problem 1

** *OK.. you notice the following alarm*
conveyer belt of pulper 1 broke down..
\- *I would.. I would stop the pulper to*
start with and then I would halt the whole
cycle afterwards and then try to repair
the conveyer belt..

...

\- *but you have to halt the whole installation,*
because otherwise they don't have any stock
anymore..

To make this point clearer one should compare this novices' protocol to that of the expert on the same problem.

Experts' protocol on problem 1

** *OK.. conveyer belt of pulper 1 broke down..*
\- *conveyer belt of pulper 1.. if that one breaks down..*
yeah.. see how long that takes to repair..
\- *not postponing the decision for very long, to*
ensure we don't have to halt the installation..

Thus, the novice does not make many elaborations on the problem statement, but immediately jumps to conclusions. In fact, he starts repairs that are not necessary at all given the situation (stopping the pulper and halting the installation). His behavior shows a lack of planning, which is a rather widespread phenomenon in novice behavior in a number of domains. For example, Jeffries, Turner, Polson, and Atwood (1981) found in the domain of designing software that experts understand a problem before breaking it into subproblems, whereas novices propose a solution before having explored aspects of a subproblem, or without showing evidence of having knowledge of relevant solutions and their applicability conditions. Schraagen (1989) found in the domain of Damage Control in the Royal Dutch Navy that novice Damage Control Officers were very much geared towards finding a solution without really exploring the problem very well in advance.

The next step in the diagnostic model, the enumeration of possible faults, also shows a difference between experts and novices as already mentioned before: the novices do not mention many faults, given the set of symptoms. This amounts to 75% less faults than experts. This is a rather striking result, since the experiment reported in chapter Four showed that novices are able to perform at a reasonable level on this topic (33% less than experts). Thus the difference between experts and novices with respect to this part of the diagnostic strategy cannot be explained by a sheer lack of knowledge about possible faults. It seems more likely that for some reason novices, although they possess this knowledge are unable to utilize this knowledge in problem-solving situations.

The next deviation from a systematic diagnostic strategy is the imbalance shown between number of faults mentioned and the number of repairs. Although not many possible faults are mentioned in certain alarm situations, novices mention just as many repairs as experts. This implies that they are not really systematic in this part of their diagnostic reasoning process, in which one would expect that an operator mentions about the same number of repairs as faults. Interesting in this behavior is that it shows some similarities with expert behavior: it seems as if the novice knows right away what the correct answer is, and what action to take, by a process of heuristic reasoning. However, since they do not have enough experience yet to have "real" and useful heuristics developed, their strategy shows signs of "pseudo-expertise" (Elshout, 1983). The pseudo-expert tries to establish immediately, by "seeing" or "knowing" what he should do or what action to take, without a careful planning of those actions, and this is what the novice operators do. For example, the repairs mentioned by novices are not very selective, or local repairs, but have a fairly general and global character, which the protocol above already showed. There is no evidence for a systematic problem analysis or diagnostic strategy, but novices come up with solutions right away.

Novices show, compared to experts, rather often signs of model-based reasoning, as is to be expected, since novices lack heuristics for this task. These models are of the following nature:

1. Reasoning based on process flow.
   For example in problem five:

   ..High level 23.. that can happen in a situation of paper breaks.. in this situation stock goes from the machine via the hog pit to 22A and from 22A to chest 23.. so with a high level of 23 you should take care that 23 does not get any supply, and therefore that 22A does not get any supply anymore..

2. Reasoning about the topographical location of parts of the process.
   For example in problem three:

   .. *Those pipes can be found just under the roof.. together with the valves..*

3. Reasoning about how the process is controlled.
   For example in problem two:

   .. *It depends whether there is a consistency indicator there.. I don't know..*

4. Reasoning about the function of parts of the installation.

For example in problem six:

.. *Since 25 is the machine chest..*

5. Reasoning about the process of paper making: the influence of process parameters on the quality of paper.
For example in problem one:

.. *If you add too much water to the stock, the paper will not get the right weight..*

However, there is only one piece of evidence for the application of a model of the process of paper making. This is not surprising, since the alarm situations in this experiment were taken from the stock preparation department and do not have a very direct influence on the quality of the paper produced.

Diagnostic behavior of the transfer subject

The transfer subject, an expert operator, but confronted with problems in a related, but different domain, uses basically the same strategy as the expert, as concluded from the data in table 5.1, table 5.2, and table 5.3. However, since he has not developed many heuristic procedures for the domain, he also uses model-based reasoning a number of times. In this respect, he does not differ much from the novices. The following protocol on problem 2 will illustrate these points. Figure 5.3 is included to present an overview of the process flow at this point.

**	*Low consistency in stock chest 2A* **	**	symptom **
-	*from 2A stock goes to the thickeners..*	**	model-based reasoning **
	let me think.. goes through the cleanerbattery.. goes through the thickeners..		
-	*no problem.. we can handle this..* *you wouldn't notice on the paper machine..*	**	judgment **
** what would you do**			
-	*see what is going on before 2A..*	**	model-based

			reasoning **
-	*maybe low consistency in the pulper..*	**	possible fault **
-	*maybe something wrong with the cleanerbattery..*	**	possible fault **
-	*all things before 2A that can cause a deviation in stock consistency.. I have to check those..*	**	possible fault **
-	*but it has no consequences for the paper machine..*	**	judgment **
-	*unless there is only water in the pulper..*	**	possible fault **
-	*then everything should be pumped back to the pulper.. and you have to start the process again..*	**	repair **

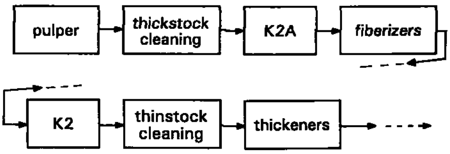


Fig. 5.3 The process flow around K2A.

It is clear from this protocol that the transfer subject does not use many heuristics when trying to find possible faults: faults are found as the outcome of model-based reasoning processes. On the other hand, he is very systematic in his reasoning process, and follows more or less the general diagnostic strategy, just like the expert. Note that he is very concerned about making a judgment about the seriousness of the problem. His priorities are set right, since his chief responsibility: keep the installation going, is always apparent. Although the transfer subject lacks specific knowledge, his reasoning process still obeys the same systematic rules as that of his expert colleague. If one would help the transfer subject in filling in details about this specific installation (system knowledge), his behavior would be like that of the expert. This is in fact a quite important result, since it gives evidence for a general diagnostic strategy as a separate layer of strategic knowledge which may be transferred to other tasks. Novices seem to have sufficient declarative background knowledge, but do not bring to bear this knowledge in a real problem, and the strategy they use is not systematic. This result is quite consistent with the results obtained by Greeno (1978) as described in the introduction to this chapter. The transfer subject uses the expert diagnostic strategy, even though he lacks specific system knowledge. This general diagnostic strategy may be a kind of long-term memory structure that is referred to by Chase and Ericsson (1981) as the "intermediate knowledge structure", and it may be very similar to what Jeffries et al. (1981) refer to as "design schema". Implications of this finding will be discussed in chapter Eight. Thus, these results give psychological evidence for the existence of strategic knowledge which is generalizable over domains and which is partly independent from system knowledge. One of the questions that can be posed, and which has not been addressed in this experiment is the interaction between strategic and system knowledge. This question could not be addressed, since scenarios are still "off-the-job" problems, in which no "real" procedures are carried out. Therefore, a second experiment was carried out to shed more light on this issue. In this study, a process simulation of part of the plant was used, in which it is possible to obtain, apart from verbal protocols, also some action from the subjects. A second goal of the new experiment was to investigate whether these results can be replicated with different subjects, to ensure some generality over different subjects. Third, since diagnosing problems on the simulator allows an operator to carry out actions, and is therefore a more naturalistic situation than troubleshooting scenarios, we will have another chance to test the ordering of categories in the model.

Chapter 6

Strategies for diagnosis: the simulator experiment

6.1 Introduction

In chapter Three a task-level model for real-life diagnosis of malfunctioning in a paper mill was presented, consisting of a number of categories together with a definition of the legitimate transitions between categories. Chapter Five described a first test of this model on the somewhat artificial task of handling troubleshooting scenarios. It was concluded that the model was not contradicted. However, this test was handicapped by the fact that, due to the experimental setting, not all categories were likely to appear in subjects' protocols. To obtain a further test of the model the experiment described in this chapter was conducted. In this experiment subjects are required to solve diagnostic problems in a context which is more naturalistic than solving scenarios, since it requires subjects to really solve problems.

6.2 Method

6.2.1 Subjects

Eleven subjects participated in the experiment, including five experts with more than five years of experience as independent operator, two trainees with less than three years of experience as operator, and who are not yet able to run the installation independently, two novices who have only been employed by the paper mill for two years, and have only received a general training on all aspects of paper making. Finally two transfer subjects participated who are expert paper makers, but who have limited practical experience in running the installation questioned about in this experiment. None of these subjects participated in the scenario-experiment reported in chapter Five.

6.2.2 Description of the simulator

To enable the observation of operators diagnosing problems in a rather naturalistic situation, a process simulation of part of the plant (the wet end and constant part), was developed in close collaboration with P. Weghorst. For a more elaborate description of the simulator, the reader is referred to Weghorst (1989). To ensure that subjects would be able to use their knowledge about the plant in the simulated environment and to give them the feeling that they were interacting with the "real installation", it was considered of the utmost importance that the simulator would have a high degree of both face validity and fidelity (Moraal, 1988). Face validity is defined as the degree of similarity between the appearance of the simulator and the real system: do the two "look alike". Fidelity has to do with the question whether the simulated system behaves similarly as the real system. The real system in this case is a process computer by means of which a paper machine is controlled. Interaction between operator and simulator takes place via computer screens showing parts of the process and a light pen for changing process parameters. Of the two types of validity, face validity turned out to be the easiest to obtain: it is possible to make the pictures shown on the screen in the simulator very similar to the pictures shown by the process computer. To obtain a high degree of fidelity a model of the installation was developed using an adapted version of Petri Nets (Peterson, 1981) as modelling theory. This model of the installation forms the basis of the simulator. The model has to be dynamic: the status of the process is liable to changes over time. To realize this aspect, procedures were developed that change the status of the model when applicable. The face validity was tested by having operators work on the simulator, and observing how easy it was for them to get acquainted with the program. Although in the simulator a mouse is used for pointing at elements on the screen, while on the real process computer a light pen is used, no operator considered this to be a problem. However, it should be noted that working with the mouse is somewhat slower than pointing with a light pen. The fidelity was tested by two independent shift managers who carefully tested the simulator on several aspects: is the process reacting in the correct manner and at the right speed to changes in controllers etc. Furthermore, if parts of the process are shut down, do the other parts behave accordingly? After a number of sessions, in between which several changes to the underlying model and its implementation were made, they decided that the simulator was behaving realistically, which was confirmed by the subjects in the experiment itself. These two independent shift managers also played a major role in determining the scenarios for the alarm situations and in indicating how these alarm situations could be solved.

From the point of view of the experimenter, the simulator has three aspects. First of all, the experimenter can define the initial status of the process: the current process values, controller settings etc. Secondly, the experimenter can determine when and what disturbances of the process occur, resulting in alarms. Finally, in the simulator all actions undertaken by the subject are registered in a log-file, enabling a more elaborate analysis of diagnostic behavior. Figure 6.1 shows an example screen of the simulator.

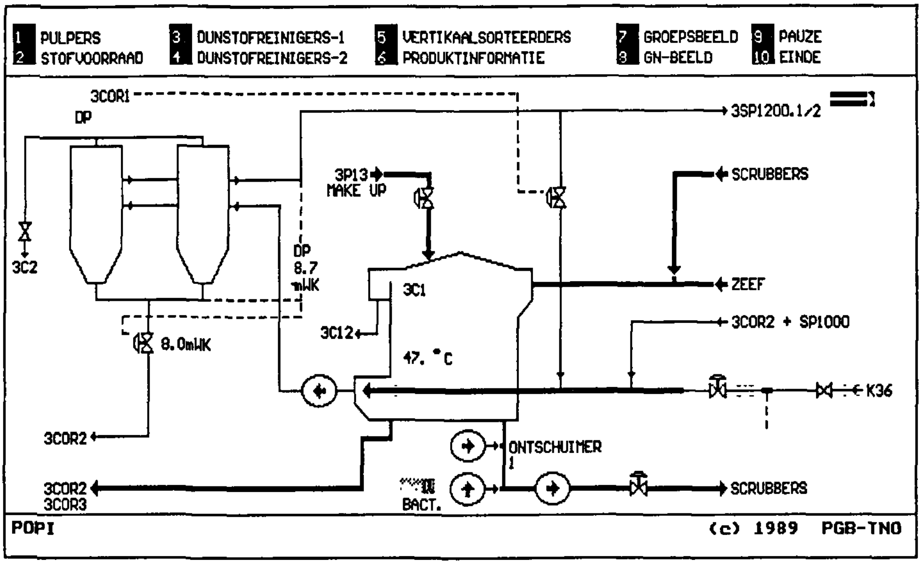


Fig. 6.1 Example screen of the simulator.

Software for the simulator was developed in Turbo Pascal 4.0 (Borland International) for a Compaq Portable 386 (20 MHz), using a separate VGA screen. Transportability of the simulator was considered important to enable experiments with the simulator in the paper mill itself.

6.2.3 Materials

Based on an analysis of possible alarms occurring in this part of the paper making process (the wet end and constant part), the following alarms were selected:

1. High level alarm chest C1
2. Wrong consistency of stock in chest K35
3. Wrong differential pressure in thinstock cleaner COR1
4. Paper break
5. Wrong differential pressure in vertical screen SP1000
6. Paper break, coupled with a suddenly occurring high level alarm of press pit C5
7. Wrong differential pressure in vertical screen SP1200
8. Wrong consistency of stock in chest K36

All selected alarm situations occur frequently in reality and were judged by an independent machine coordinator as being not too difficult to solve and as representing a wide range of complexity of diagnosis for an operator. The alarm situations are quite different in nature from each other. For example, a wrong differential pressure in thinstock cleaner COR1 within the limits presented in this experiment (problem 3) requires an operator simply to correct it, without performing a thorough analysis. On the other hand, a wrong differential pressure in the vertical screens (problems 5 and 7) should lead to a thorough analysis of its causes and the way it influences other parts later on in the process. When paper breaks occur (problems 4 and 6), the first thing to do would not be a thorough analysis of its causes, but checking quickly whether the installation reacts appropriately, for instance whether the water supply to, in this case, the press pit is running and sufficient, and whether both agitators are running. All these actions are aimed at preventing the press pit to overflow, which, when it happens, will certainly lead to an enormous mess, but may also damage the pickup felt due to the enormous amount of paper pushing against the felt, at least when one forgets to raise this felt quickly enough in this situation.

The alarm situations also differ in the ease with which the causes for the alarms can be spotted. In some cases, for example alarm situations 1 and 7, the causes are rather easy to find, mainly because the fault can be read of the same screen alerting to the alarm situation. An example of this would be a wrong setpoint for a controller. In other cases (problem 4 and 6), the operator has to do a lot more reasoning to find a probable cause for these alarms, and has to act according to what he thinks the fault is or has been, since the fault will not show up on the screen (for example clogging of a vertical screen will never show up, but may be inferred from process values).

Finally, in problem 5 the fault behind the alarm is extremely hard to spot, since one should be able to read it off the screen, but the process computer "lies", e.g. sends a wrong signal. In this problem one of the agitators does not start, which does not show up on the screen, and can only be deduced after checking all the other possible sources for this alarm and sending someone to the agitator to check whether this machinery is actually running improperly.

6.2.4 Procedure

The experiment consisted of two sessions, the first session geared towards getting the subjects acquainted with the simulator. The second session was the experimental session in which subjects had to diagnose the alarms. Both sessions took place either in a quiet location remote from the paper machine or in the control room, depending on whether the operator could be dismissed for the moment from his duties as supervisor. This implied that he did not have to do any other tasks apart from the experiment, unless some serious trouble occurred, in which case the experiment was halted for the moment. If the experiment was run in the control room, care was taken to ensure that no interference with the real situation would take place. This went quite well, apart from one occasion, in which paper breaks occurred at almost the same time in the simulator and on the real paper machine. In almost all cases, the experiment took place during normal working hours.

In the first session, subjects were first given instructions by the experimenter about the scope of the simulator and how it had to be controlled. After this, they were allowed to play with the simulator as long as they needed to feel confident that they knew how to control it. This first session was also aimed at giving subjects confidence with regard to the validity of the simulator: was it really behaving as the real installation. When subjects indicated that they knew how to control it, their knowledge was tested by asking them to carry out certain actions on the simulator, such as switching between screens, stopping and starting a pump, changing a setpoint value, switching from automated control to manual control and vice versa, and finally getting information about a certain part by asking an "imaginary colleague" to check the suspicious part. Most of the subjects indicated spontaneously that they felt the simulator was very much like the real paper machine in its behavior. All subjects finished the first session within 45 minutes.

The second session, which was held one or two days after the first session, started with a short reintroduction of the simulator, after which the subject was given some time to get himself reacquainted with the simulator. When he indicated he was ready to start the experiment, he was given the following instruction: "Imagine you're the paper maker. In what follows a number of alarm situations will occur, one at a time. I would like you to solve these alarm situations while thinking aloud. By thinking aloud I mean that you tell me as much and as exactly as possible what you are thinking about and do in this situation." After this, the first alarm situation was introduced, which was treated as a practice alarm and was not taken into account in the analysis. Following the practice problem the subject was given feedback about his performance, after which the experiment started. The whole session was recorded on tape, for which the subject had given his permission in advance of the experiment. Subjects were told that the results would be regarded as confidential: no information about the performance of a specific subject would be revealed to anyone else.

All subjects received all alarm situations in the same order, and they were told by the experimenter that they could use as much time as they needed to solve the problems. All of them finished the session within one hour.

6.3 Results and discussion

All protocols were transcribed verbatim and, based on pauses in speech, divided into statements or idea units. After the transcription, the corresponding part of the log-file of the simulator was put into the protocols on the appropriate places, mostly determined by listening where on the tape keyboard clicks occurred, sometimes determined by a synchronization of the verbal protocol on tape and the log-file of the simulator, and in a few remaining cases determined by guessing. The protocols were scored according to the categories of the general diagnostic model as presented in chapter Four. One extra category was added, "comments", which was used for task-irrelevant comments by the subject. The scoring was done independently by two persons, one being the experimenter, and the other one a psychologist with sufficient knowledge of paper making, and therefore acquainted with the concepts used by the subjects. This second person was not aware beforehand whether a certain protocol belonged to a novice, a trainee, an expert, or a transfer subject. Even without discussion there was a high level of agreement (82%) upon the assignment of categories to statements in the protocols. It turned out that 90%

of the statements could be assigned to categories of the diagnostic model, which gives confidence in the sufficiency of the model for the purpose of this study.

In the following, the results of this experiment will be presented. Special attention will be given to the questions posed in the previous chapter: can the results described in chapter Five be replicated with different subjects and in a different situation which was more adapted to reality.

Table 6.1 gives an overview of the results from this experiment with respect to the frequency of occurrence of the categories Judgment, Evaluation, Consequences as extracted from the protocols, summarized over all 7 alarm situations, together with the overall number of categories of the diagnostic model used by the different groups of subjects. The results for each group have been averaged over subjects.

Table 6.1 *Categories from the diagnostic model, summarized over all 7 alarm situations, for experts, transfer subjects, trainees and novices. The results for each group have been averaged over subjects.*

	Experts (n = 5	Transfer (n = 2)	Trainees (n = 2)	Novices (n = 2)
Judgment (max = 7)	5.6	4	1.5	0
Evaluation (max = 7)	5.4	6	2.5	2
Consequences (max = 7)	4	5.5	1.5	1.5
Number of categories used, summarized over all 7 problems (max = 56)	39.8	43.5	28.5	26

A Chi-square on the total number of categories used reached significance, Chi-square = 79.26, df=3, p <.001, implying that there is a difference in number of categories used between the various groups, with the experts and transfer subjects using more categories than the trainees and novices. This difference was more thoroughly investigated by computing Chi-squares on those categories of which only the occurrence in a diagnostic protocol, and not so much the number of times that the category occurs in a protocol is of importance, which are the categories Judgment, Evaluation, and Consequences. A Chi-square on those categories revealed significant differences over groups. Chi-square on Judgment = 31.45, df=3, p <.001, Chi-square on Evaluation =

16.15, df=3, p <.01, Chi-square on Consequences = 14.33, df=3, p <.01. These results should be taken as evidence for a difference with respect to the usage of the categories Judgment, Evaluation and Consequences between the experts and transfer subjects on the one hand, and the trainees and novices on the other hand. Interestingly, the category "Ordering of Faults", that did not appear in the protocols obtained in the scenario-experiment, also did not show up in this experiment. Thus, it appears a rather stable, though puzzling phenomenon, since one would certainly assume that subjects make some sort of ordering of faults before continuing. It may be that this is a very implicit process which does not show up in verbal protocols. The only evidence for "Ordering of Faults" was actually obtained after an alarm situation had been solved, since the subject stated the following as a comment on problem 6:

- *That is the last thing you would look at.. the agitators..*
- *when the level is raising the first thing to do would be looking at the water valve..*
- *If those are allright.. you go to the pumps..*
- *The agitators are the last ones.. and especially when on the screen you see them running..*

However, since this statement came after the problem had been solved and the computer was loading the next alarm it was not taken into account in the analysis and treated as a comment.

Table 6.2 gives an overview of the performance of the subjects on the categories Determination of Faults (number), Tests (number), and Determination of Repairs (number).

Table 6.2 *Performance on the categories Determination of Faults, Tests, and Determination of Repairs, summarized over all 7 alarm situations, for experts, transfer subjects, trainees and novices. The results for each group have been averaged over subjects.*

	Experts (n = 5)	Transfer (n = 2)	Trainees (n = 2)	Novices (n = 2)
Number of Faults	24	37.5	8	13
Number of Tests	20	34.5	6	10.5
Number of Repairs	15.8	12.5	10.5	10

The reader should notice that the number of tests can never be greater than the number of faults, simply because one can only test those faults that are taken into consideration.

The Kruskal-Wallis Test Statistic T revealed marginally significant differences between the different groups and number of faults mentioned, $T = 6.911$, $p = .07$, and between the various groups and number of tests mentioned, $T = 7.26$, $p = .06$. However, no significant overall difference was found between the four groups and number of repairs mentioned. These results imply that there appears to be a difference, although only marginally significant, between experts and transfer subjects on the one hand and trainees and novices on the other hand.

These tables show that the results presented in chapter Five could indeed be replicated, since the same pattern of results emerges from both studies. In chapter Five, it was concluded that there are differences between expert and transfer subjects on the one hand, and the novices on the other hand with respect to the number of faults mentioned. Again, this result shows up in this experiment, with the trainees behaving like novices. In the experiment presented here the category "Judgment" is lacking in the protocols from trainees and novices which is a very similar result to the results obtained in the scenario experiment. Usage of the category Evaluation could not be tested in the previous experiment, due to the experimental situation, but the experiment reported here showed that trainees and novices do not use it very often compared to experts and transfer subject. This result is therefore an extension to the results obtained in the previous experiment. Finally, the overall quality of the diagnostic reasoning process, as measured by the number of categories from the diagnostic model used, shows the same pattern as the results from the previous experiment, although the results may not be as extreme as presented before.

Chapter Five suggested that inexperienced subjects show signs of pseudo-expertise (Elshout, 1983), that is, they try to establish immediately what to do or what action to take, without a careful planning of those actions. The results presented here suggest an extension to this hypothesis, since one of the things lacking in novices' protocols is a systematic evaluation afterwards whether the actions undertaken result in desired effects. They appear to be so certain about the results that it almost seems as if no check afterwards is necessary. The following excerpt from a novice' protocol on problem 5 (wrong differential pressure in vertical screen SP1000) will illustrate this point. This protocol is a combination of the verbal protocol delivered by the subject mixed with the appropriate pieces of the log-file of the simulator. The pieces of log-file have a time stamp on them together with the time in seconds passed since the problem was started. Explanation for the reader is put between **.

- *let's see where..*

11:38:26 90 to screen: VERTICAL SCREEN
11:38:43 107......................Action performed on P2

-	*vertical screens..*	** place of symptom **
-	*it is raising..*	** differential pressure **
-	*so I will lower it somewhat.. pump C2..*	** flow is lowered from C2 to SP1000
-	*so it will not send too much through it..*	** lowering flow by controlling pump C2 **
-	*let's see where it goes to..*	** model-based reasoning **
-	*reject sorter..*	
-	*I will add some water.. if it gets clogged..*	** the vertical screen **
-	*If I will stop that pump I will also stop that vertical screen..*	

11:38:53 117 clicking on status of P2
11:39:00 124......................Action performed on P2

11:39:13 137 clicking on AUTO
11:39:16 140 clicking on DO IT ** subject puts pump P2 in automatic mode **
11:39:16 140 P2 in automatic mode
11:39:27 151......................Action performed on P2
11:40:05 189......................Action performed on P2
11:40:06 190 No system actions anymore

- *in fact it doesn't make any difference* ** putting the pump in automatic mode **
- *I will simply turn it off..* ** P2 **

11:41:07 251 clicking on status of P2
11:41:14 258 clicking on HAND
11:41:16 260 clicking on TURN OFF
11:41:18 262 clicking on DO IT
11:41:18 262 P2 turned off
11:41:25 269....alarm SP1000 cancelled ....

- *problem solved..*

Figure 6.2 gives an illustration of the process flow at this point.

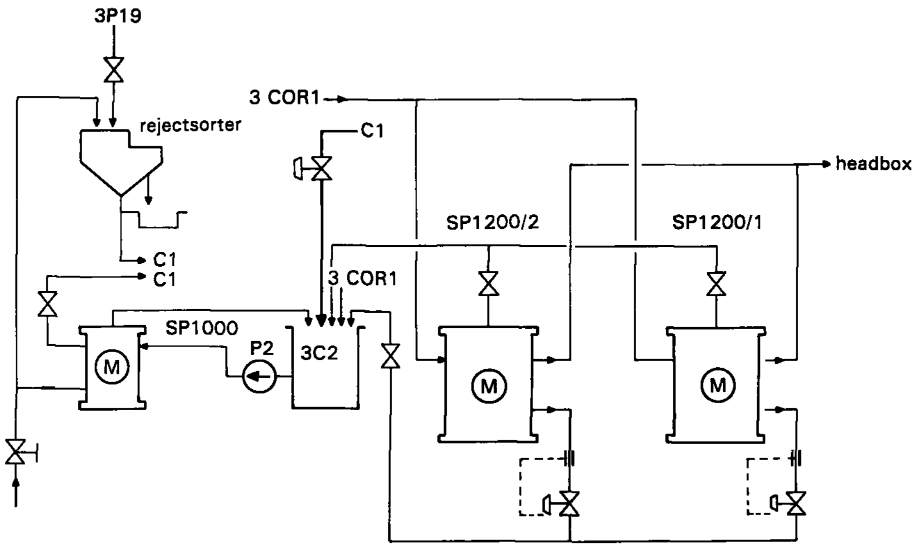


Fig. 6.2 The process flow around water chest C2 and SP1000 .

The differential pressure in vertical screen SP1000 gets too high, which implies that either the input pressure (coming from C2) is too high, or the output pressure (going to the reject sorter) is too low. What the subject should do is judging what the implications are of this alarm for the production process, followed by an analysis of its possible faults (for example a clogging of the vertical screen). A good repair would be a further opening of the valve between C1 and C2 resulting in an overflow of C2 which will lead to a removal of dirt, or lowering the amount of reject. These repairs are described in the expert protocol on page 106 onwards. What the subject in fact does is turning off pump P2, resulting in a zero differential pressure in the vertical screen, which of course leads to a cancellation of the alarm, but is obviously no solution to the problem. This behavior is quite comparable to turning off the engine of your car when warning signs (for example the oil pressure gauge) appear on your dashboard and then drawing the conclusion that there is no problem left.

In the previous chapter it was concluded that the model for diagnosis as presented in chapter Four was not contradicted by the empirical data. Those deviations that were found can be traced to instantiations of depth-first reasoning, in which a person fully completes his line of argumentation for a certain fault, including the repairs and consequences, before taking up another possible fault. The model however, assumed that subjects would follow the categories one by one, thus completing the reasoning process about one category before taking up another one. It might be, though, that those deviations of the model are caused by the experimental situation and the somewhat artificial task of solving scenarios.

Two categories of the model that almost never showed up were "Evaluation" and "Ordering of possible Faults in terms of likelihood". An explanation for the absence of "Evaluation" was given before, but one may wonder whether the apparent lack of "Ordering" is a stable phenomenon. As described before, "Ordering of possible Faults" again does not show up in this experiment, presumably because it is a very implicit process that does not show up in verbal protocols.

A further aim of the experiment reported here is to see whether the results about the ordering of categories of the model could be replicated with different subjects in a more naturalistic situation. In the following an analysis of the transitions in the protocols of expert operators, transfer subjects, trainees and novices will be presented, coupled with a more elaborate analysis of the results of each group. This will be followed by a discussion of the types of device knowledge used by subjects insofar evidence for it could be found in the verbal protocols.

Expert subjects

Table 6.3 gives an overview of the transitions in the protocols of expert subjects.

Table 6.3 *Transitions in the protocols of expert subjects (n = 5). Legitimate transitions are typed in italics. The transitions have been summarized over subjects and alarm situations.*

from				to				
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*16*	*7*			*10*		
Judgment			*13*			*9*	*1*	*1*
Faults					*62*	*7*		
Ordering								
Testing		*2*	*60*	*1*		*21*		
Repairs		*5*	3		*2*		*32*	*6*
Conseq.		*1*	4			*14*		*13*
Eval.		*1*				*1*		

Out of a total number of transitions of 292, only 7 (2%) turned out to be illegitimate transitions, falling into two types: Repair --> Fault, and Consequence --> Fault. When looking at the protocols, the illegitimate transitions can be explained in the same way as they could in the previous study: what experts do is reasoning in a depth-first way: first complete your line of reasoning about one possible fault, including repairs and consequences, before taking up another one. Presumably, this depth-first reasoning has to do with preventing the occurrence of working memory overload. This forms a replication of the results presented in chapter Five.

An interesting point to note in the protocols of two of the expert subjects is the occurrence of a new category that has not been defined in the diagnostic model as described in chapter Four. What they do in terms of repairs is not only a determination of the consequences of the repairs, but also an *ordering of repairs* in terms of desirability of the consequences of each repair, which obviously is only possible when more than one repair is taken into account. The following protocol excerpt on problem 5 illustrates this point. This excerpt contains, for reasons of clarity, only the handling of repairs by the subject.

- *The pressure is too high..*

...

- *We could do a number of things..*
- *By a further opening of the valve between C1* ** first repair with highest preference **
  and C2 we may try to get rid of the dirt in C1,
  because an overflow of C2 will lead to a removal
  of the dirt towards the drain..
- *that is done quite often.. let's try that..*

...

- *other possibility that is seen often is..*
- *much faster and more effective,*
- *is switching pump 2 on and off..* ** second repair, implies some risks, is therefore not applied right away **
- *but that leads to a swing in thickstock supply..*
  and there is a chance that a paper break occurs..
- *another possibility is.. since the differential*
  pressure in the SP1200 is reasonable..
  we could somewhat lower the amount of reject.. ** third repair with second preference **

** subject undertakes this repair **

...

** subject switches pump 2 off and on **

- *the last thing to do would be changing the* ** fourth repair, lowest preference **

  inlet and outlet somewhat.. to bring the differential pressure back within tolerance..
- *but that is the last thing that you would do..*

Transfer subjects

Table 6.4 gives an overview of the transitions in the protocols of transfer subjects:

Table 6.4 *Transitions in the protocols of transfer subjects (n = 2). Legitimate transitions are typed in italics. The transitions have been summarized over subjects and alarm situations.*

from	to							
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt		*2*	*5*			*4*		
Judgment			*7*			*1*	*1*	
Faults		*1*			*30*	*4*		
Ordering								
Testing			*27*			*12*		
Repairs			3		*1*		*12*	*3*
Conseq.			2		1	*4*		*5*
Eval.			2			3		

Out of a total number of transitions of 127, 11 (9%) could be considered illegitimate. They are of the following types: Repair --> Fault (3), Consequence --> Fault (2), Evaluation --> Fault (2), Consequence --> Test (1), and Evaluation --> Repair (3). The first two types (Repair --> Fault and Consequence --> Fault) can be taken as evidence for depth-first reasoning as explained before. The transition

Evaluation --> Fault should be taken as evidence for a startup of a new search process. The subject has pursued a certain hypothesis (possible fault), but realizes that the hypothesis is wrong since the repair does not alleviate the problem, and he therefore starts with a new hypothesis. The following protocol excerpt on problem 2 (wrong consistency of stock in chest K35) will clarify this point. Figure 6.3 gives an illustration of the process flow around K35.

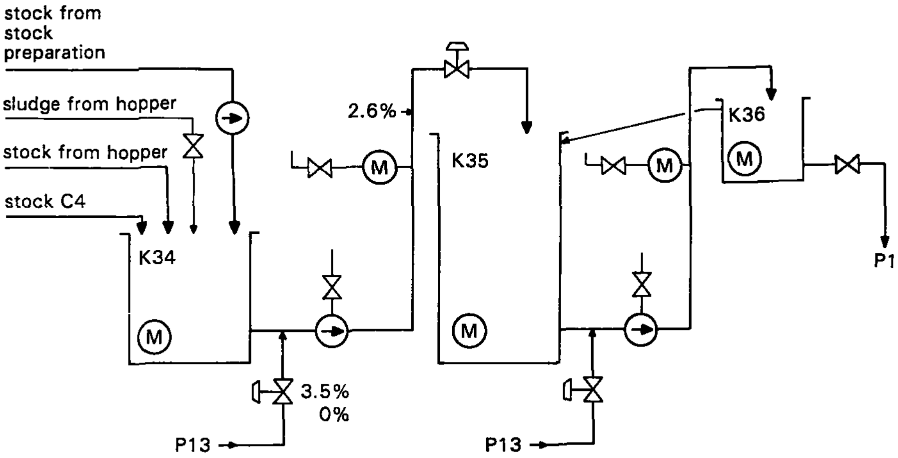


Fig. 6.3 The process flow around K35.

.... ** the protocol excerpt starts
in the middle of the problem **

- *I have a low consistency in chest K34 of 2.6 ..*
- *That implies that either water*
  is supplied in K34 ..
- *Or somewhere before..* ** from stock preparation **
- *I do not get any stock from stock*

	preparation right now..	
-	*so that cannot be it..*	
-	*I don't see anything wrong with C4..*	
-	*I think there is a problem with the consistency controller of chest K34..*	** fault **
-	*Let me send someone over there to check whether he is returning the right value..*	** testing **

```
13:48:18  760 Information asked about CONSISTENCY CONTROLLER K34
13:48:57  799 Information given about CONSISTENCY CONTROLLER K34
```

-	*No problem found.. so that one is correct..*	** test result: no problem found **
-	*This will not work.. I still have that problem..*	** evaluation: problem not solved
-	*So I have got to get back to the pulpers.. there should be the cause..*	** new hypothesis **

The transition Consequence --> Test can only be explained by assuming that in fact it is a transition Consequence --> Fault --> Test, with the fault not being mentioned, implying that it is another indication of depth-first reasoning.

The transition Evaluation --> Repair is an interesting one, since it appears that apart from starting up a new hypothesis when the evaluation turns out negative, it is possible to try a new repair, or refining a certain repair. This process should be regarded as adjusting a certain process value a little bit, evaluate the outcome, adjust a little more, and so on. The following piece of protocol illustrates this point:

- *Let's adjust grammage..*
- *By supplying thickstock via GN-screen..*

```
13:38:16    158   to screen GNSCREEN
13:38:28    170   clicking on setpoint controller chest K36-C1
13:38:42    184   setpoint changed to 1000.00
```

- *200 liter thickstock added..* ** repair **
- *what's the effect on grammage..*
- *200 liter is not sufficient..* ** evaluation **
- *I will gradually increase the value..* ** repair **

13:39:35 237 clicking on setpoint controller chest K36-C1
13:39:40 242 setpoint changed to 1200.00

- *200 liters added..*
- *the grammage is coming up..*

.....

As found in the previous experiment, whenever Symptom --> Repair transitions occur in both the protocols of experts and transfer subject, they are always accompanied by a judgment that the problem is in fact a serious one, in which it is important to take immediate actions.

Trainees

Table 6.5 gives an overview of the transitions in the protocols of trainees.

Table 6.5 *Transitions in the protocols of trainees (n = 2). Legitimate transitions are typed in italics. The transitions have been summarized over subjects and alarm situations.*

from	to Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt			*8*			*7*		
Judgment			*3*					
Faults					*11*	*2*		
Ordering								
Testing	**1**		*3*			*6*	*1*	
Repairs		*1*	**2**		*1*		*1*	*4*
Conseq.			**1**					
Eval.								

Out of a total number of transitions of 52, only 4 (8%) could be considered illegitimate. Three of them are "known types", Repair --> Fault and Consequence --> Fault have been explained before as instances of depth-first reasoning. The final one, Testing --> Symptom is an odd one, that cannot be easily explained. By inspecting the protocol it emerged at an occasion in which the subject only gradually becomes aware of the nature of the symptom:

19:47:15 73 ALARM..
19:47:28 86 To screen: STOCK SUPPLY

- *So it is here.. let's see.. ooh.. there..*

19:47:50 *108* *To screen: PULPERS*

** What are you looking at? **
- *Pulpers.. whether a valve is wrongly opened.. thinning water..* ** Fault **
- *it doesn't look like it..* ** Testing **
- *Let's see..*
- *There is nothing wrong....*

.....

- *setpoint of 70% on C2 is allright..* ** Fault/Testing **
- *Look at the GN screen..*
- *not much to see..*
- *So the consistency is too thin..* ** Symptom **

** What are you looking at **
- *It doesn't have to do anything with this..*
- *I don't know..*

Novices

Finally, the results of the novices will be discussed. Table 6.6 gives an overview of the transitions in the novices' protocols.

Table 6.6 *Transitions in the protocols of novices (n = 2). Legitimate transitions are typed in italics. The transitions have been summarized over subjects and alarm situations.*

from	to							
	Sympt	Judgm.	Faults	Order.	Test	Rep.	Cons.	Eval.
Sympt			*6*		1	*8*		
Judgment								
Faults					*13*	*5*		
Ordering								
Testing			*13*			*4*		
Repairs			5		*1*		*5*	*2*
Conseq.					1	*2*		
Eval.								

Out of a total number of transitions of 66, 7 (11%) were considered illegal. Five of them are of the type Repair --> Fault and have been described before. One of them, Symptom --> Test presumably should be Symptom --> Fault --> Test, with the fault not being mentioned in the protocol. The transition Consequence --> Test was discussed before and explained as a transition Consequence --> Fault --> Test, turning it into another occasion of depth-first reasoning.

Interestingly, in this experiment as well as in the previous one, we find that Symptom --> Repair transitions in both the protocols by trainees and novices are **not** accompanied by a judgment about the seriousness of the problem. It has been argued in chapter Five that this is in fact a rather important judgment to make, which in expert behavior may change the diagnostic strategy.

Types of system knowledge used by subjects

In chapter Five an overview was given of the different types of system knowledge that is used by subjects coupled with an analysis of the number of times a subject has to use system knowledge explicitly to continue his reasoning process. By means of model-based reasoning the subject finds a way to resolve the existing impasse. It was concluded that novices and transfer subjects had to rely much more on model-based reasoning than experts. Table 6.7 gives an overview of the results obtained in this experiment. Only those instances of model-based reasoning that contribute to the resolving of an impasse are counted, other instances of the use of system knowledge are left out.

Table 6.7 *Instances of model-based reasoning for experts, transfer subjects, trainees, and novices. The results have been averaged over subjects.*

	Experts (n = 5)	Transfer (n = 2)	Trainees (n = 2)	Novices (n = 2)
Model-based reasoning	1	5	5	6

Since the main interest in the use of model-based reasoning concerns the different amount of usage of model-based reasoning between the experts on one hand, and the three other groups on the other hand, an analysis was carried out accordingly. In this analysis, the experts are taken as one group, and the transfer subjects, trainees and novices together are taken as the other group. This analysis resulted in a significant difference between the two groups, Mann-Whitney U = 30.00, $p < .01$, which implies that experts make less use of model-based reasoning than subjects with less experience in the specific domain, which is in line with the general opinion about this matter.

Apart from the question how many times different groups of subjects have to rely on model-based reasoning to complete their diagnostic process, a more interesting question concerns the *types* of system knowledge used by subjects, to get an indication of the necessary system knowledge to perform diagnostic tasks in this situation. In chapter Five, the following types of system knowledge were identified, which were also present in the protocols in this experiment.

- Knowledge about the process flow.
  For example:

  ".. Stock flows from C5 to C4, and from C4 it goes to K34.."

- Knowledge about the topographical location of parts of the process.
  For example:

  ".. The amount of reject cannot be controlled from the operator's cabin, you have to change it manually on the reject sorter itself.."

- Knowledge about how the process is controlled.
  For example:

  "..In normal situations this valve is always put in manual mode.."

- Knowledge about the function of parts of the installation.
  For example:

  ".. This valve controls the consistency. If the consistency gets too low it should be opened some more.."

- Knowledge about the process of paper making: the influence of process parameters on the quality of paper.
  For example:

  ".. Feeding back water into the system should be done very gradually, otherwise oscillations in grammage will occur.."

However, apart from the types described above, some additional ones could be identified:

- Knowledge about normal values of process parameters.
  For example:

  ".. The differential pressure in the thinstockcleaners is usually around 10.."
  ".. A value of 218 is no problem.."

- Knowledge about process dynamics.
  For example:

  ".. This valve is reacting too slowly.."
  ".. With the valve opened this far, the differential pressure should already have been higher.."

- Knowledge about the functioning of parts of the installation. This should not be confused with the *function* of a certain part. The *function* of a certain part is for example cleaning or refining. The *functioning* however, has to do with the way the function is realized in a global technical sense.
  For example:

  ".. The differential pressure is determined by the difference in pressure between inlet and outlet. If the inlet remains the same, then a lowering in differential pressure can only be caused by a problem on the outlet side.."

6.4 General discussion

The experiment reported in this chapter aimed at a further test of the model for diagnosis as described in chapter Four. Do subjects use the categories as described in the model, and is their strategy in accordance with the predicted transitions between categories? In general the data of the experiment described in chapter Five and the experiment described here are quite consistent with each other and with the model, thus giving some confidence in the model developed.

The results of both experiments as a whole showed that some revisions to the model are needed. Firstly, it appears that one diagnostic category was lacking from the model: ordering of repairs. Secondly, the model as initially described in chapter Four assumed a much more breadth-first approach than subjects actually undertake. The model assumed that the different categories, which are in fact taken as subgoals of the whole task of diagnosis, would subsequently be pursued by the subjects in a serial fashion. However, it turned out that subjects proceeded in a much more depth-first manner. They start with a possible fault for a symptom and proceed through the whole reasoning process, including reasoning about repairs and consequences, until sufficient evidence has been obtained to either discard the likelihood of this fault as being the present fault, or concluding that this must have been the present fault since the problem could be solved after application of appropriate (local) repairs.

This leads us to the following expert flow-chart model for diagnosis as expressed in figure 6.3. With regard to the diagnostic strategy undertaken, no systematic differences were found between the experts and the transfer subjects, suggesting that even if system knowledge is not sufficiently available an operator may still be able to follow a systematic strategy. Therefore, the model postulated is a model for the diagnostic strategy employed by expert operators in general.

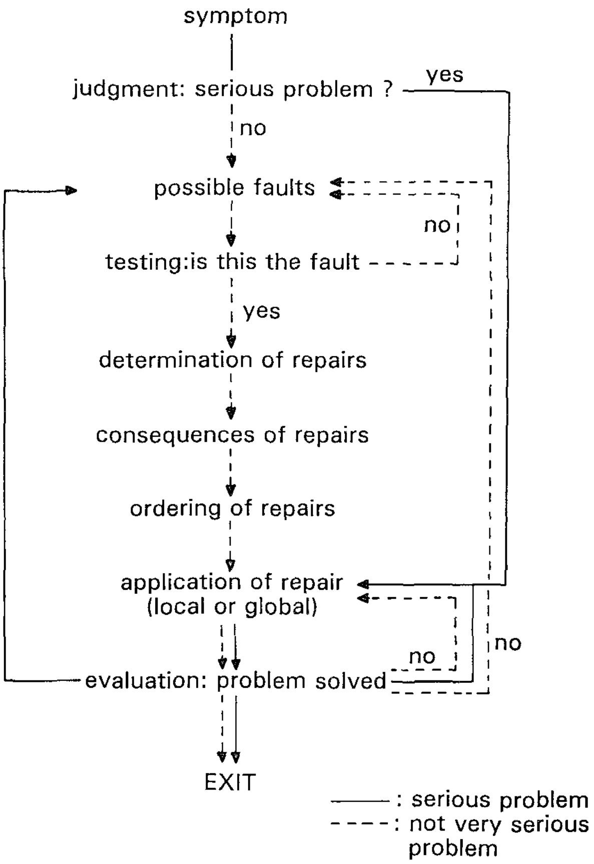


Fig. 6.4 A flow-chart model of the diagnostic strategy applied by expert operators.

This model shows the different strategies an expert operator may follow. If the problem is judged to be serious, the operator will immediately continue with the application of a (global) repair, followed by an evaluation whether the problem has been solved. This process may be followed by a more thorough diagnosis in which the fault causing the alarm is diagnosed in order to determine the correct local repair, ensuring a solution "once and for all". If the problem is not a very serious one, the subject will consider possible faults one by one and test them, until a likely one is found. This is then followed by a determination of repairs, their consequences, an ordering of repairs (if necessary), application of repairs and an evaluation whether the problem has been solved. If not, the expert might do two things: either try another repair, or back up higher in the tree: he may realize that he has not yet spotted the actual fault, and therefore the problem has not been solved. In case no possible faults are left, or the operator cannot think of any other faults than the ones he already tested, he will be inclined to use a global repair to alleviate the problem.

The model describing the diagnostic strategy employed by novices and trainees (inexperienced operators) differs from the expert model and is far more simple. Firstly, some of the categories (Judgment, Evaluation and Consequences) are lacking altogether. Also, novices jump much more quickly to repairs, without realizing whether a certain repair actually is right for a certain situation.

Figure 6.5 gives an overview of a model for the diagnostic strategy employed by inexperienced operators.

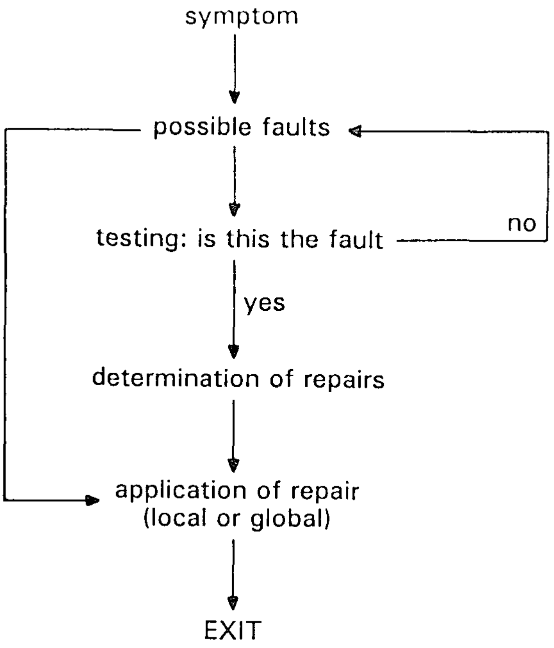


Fig. 6.5 A flow-chart model of the diagnostic strategy applied by inexperienced operators.

One of the most important results of this chapter and the previous one is the finding that transfer subjects, although they do not have sufficient system knowledge, are still able to employ a systematic diagnostic strategy that is not very different from the strategy employed by experts. This gives evidence for the existence of a *task-dependent* strategy for diagnosis at least in this type of technical domain, independent of the availability of system knowledge. Implications of these results for the acquisition of diagnostic skill will be discussed in chapter Eight. Before that, it is important to have a closer look at the system knowledge used by subjects. This chapter provided some suggestions about the nature of this system knowledge.

The question remains, though, how the different types of system knowledge used as identified in this chapter and the previous one may be integrated into the model for diagnostic reasoning. Which types of system knowledge are being used in each phase of the diagnostic process? For the diagnostic situations as reported in this study, this results in the following table:

Table 6.8 *Relation between phases of the model and used system knowledge. 1 = Symptoms, 2 = Judgment, 3 = Possible faults, 4 = Ordering of faults, 5 = Testing, 6 = Determination of repairs, 7 = Consequences of application of repairs, 8 = Evaluation.*

	1	2	3	4	5	6	7	8
Process flow		*	*		*		*	
Top. location						*		
Controls	*		*		*	*	*	
Function comp.	*	*	*			*	*	
Paper making	*	*	*	*		*	*	*
Normal values	*		*		*			*
Process dynamics	*		*		*			*
Functioning comp.			*			*		

Since the types of knowledge identified are obtained in experiments with operators as subjects, one should keep in mind that this table gives an operator-oriented point of view, which may be rather different from a technicians' point of view (Schaafstal, 1989). To gain insight in the availability of different types of system knowledge by expert and novice operators, compared to system knowledge of technicians, a number of experiments was conducted, which are described in chapter Seven.

Chapter 7

The availability of system knowledge

7.1 Introduction

In the previous two chapters, results were discussed of two experiments concerning the type of diagnostic strategy used by experts and novices. Evidence was found that the application of a diagnostic strategy is somewhat independent of the domain or system knowledge available. First, transfer subjects who lack specific domain knowledge are still able to solve problems in a very systematic way, similar to expert operators (chapters Five and Six). Second, although novices may have domain knowledge available when explicitly asked about (chapter Four), they do not utilize this knowledge very well in problem solving tasks.

However, no one would argue that a good strategy alone makes a perfect diagnoser. A diagnostic strategy should be treated as a control structure, or psychologically speaking as metacognitive knowledge, and is in fact an empty shell, which has to be filled with knowledge about the specific problem domain and problem situation. For very well known situations heuristics (immediate links between conditions and actions) may be developed that bypass the use of system knowledge, but when someone is confronted with a novel situation, he or she has to fall back on necessary system knowledge to deduce the correct answer. In this situation, a good diagnostic strategy will help a person decide what to search for, and will keep him/her on the right problem solving track. Clearly, there are both a large amount and many types of knowledge that a person can have about the system, varying from the name of the installation to its functions and purposes and how it works (see Kieras, 1987). Quite some research has been carried out about the value of system knowledge or, in a broader sense, mental models. For example, Kieras and Bovair (1984) conducted a series of experiments on a simple control panel system, in which the user's task was to route power through a series of components and controls to a specified destination. Indicator lights showed which components were operating properly, and where power was in the system depending on the settings and controls. Compared to groups that did not know the device topology, subjects with adequate system knowledge were able to learn how to operate the device more quickly and were able to retain the procedures much better as well. In addition, subjects with adequate system knowledge were able to infer very readily how to operate the device, compared to those who did not have this knowledge. Interestingly, it was specifically the system topology, such as schematic diagrams, that was important for effective system knowledge. Generalized knowledge about how the system and its components worked was of no value. The work by Halasz and Moran (1983) on calculators provides an illustration of how the value of system knowledge depends on the task that a user must do. Using a reverse-Polish calculator, Halasz and Moran found that subjects who understood the internal stack architecture of the calculator were able to make use of this knowledge in efficiently devising solutions to problems that could indeed benefit from use of the stack. However, simpler problems that could be solved routinely did not benefit from such knowledge, presumably because the user did not have to do any substantial inference in order to solve them. Results obtained from studies geared towards the training of troubleshooting or diagnostic skills indicate that instruction in theoretical principles (an element of system knowledge) is not an effective way to produce good troubleshooters (Morris and Rouse, 1985). Although these results are obtained in studies on electronic troubleshooting, the results are consistent with findings from other domains such as process control (Brigham and Laios, 1975; Crossman and Cooke, 1974; Kragt and Landeweerd, 1974; Morris and Rouse, 1985) and mathematical problem solving (Mayer and Greeno, 1972; Mayer, Stiehl, and Greeno, 1975), in which explicit training in theories, fundamentals, or principles failed to enhance performance and sometimes actually degraded performance. Morris and Rouse (1985) therefore conclude that "fundamental understanding has been empirically shown to be useful for answering theoretical questions but not for solving problems." The previous two chapters already gave some hints to what types of system knowledge are used by operators when diagnosing problems. System knowledge is defined very broadly: all kinds of knowledge about the installation used in problem solving. Bainbridge (1988) puts forward a list of five main types of knowledge used by operators in general. That list is a simplified version of Rasmussen's (1979) knowledge 'models'. The following distinctions are made, which will be used as a framework to classify the types of knowledge used by operators in diagnostic tasks as discussed before:

1. A definition of the product, and the sequence of transformations the input materials go through resulting in the product.

   Knowledge about paper making fits into this category, together with *knowledge about the process flow*. Knowledge about the process flow is actually a hard one to classify, since in the paper mill in this study it is related to both the sequence of transformations the input materials go through to make the final product, and to process functions and behavior.

2. The process limits, and the constraints on its operation. These limits have to do with the physical configuration and behavior of the process that are needed to operate safely and efficiently. These limits may change according to the process phase. Efficiency criteria for instance, are less important during start-up or fault management than during production, whilst process limits are especially important in batch processes which consist of a series of stages.

   Knowledge about the normal values of process parameters belongs to this type of knowledge.

3. The process functions and behavior, which includes the functions of the process, such as cooling or filtering, and the cause-effect (cause-consequence) relationships between them, the process variables within each function, and the cause-effect relationships between those variables, and the parameters of each of these cause-effect relationships: its gain and time constants, and the amplitude and variability of change.

   In the protocols the following instances of this type could be observed: *knowledge about the function of process components*, *knowledge about process flow*, and *knowledge about process dynamics*.

4. The plant's physical structure: how the functions are realized. This can be divided into a number of aspects, among which the major physical groupings of components, the probability of failure, the spatial position, and the question whether a physical component can change state (e.g. a valve can be either opened or closed).

   In the protocols the following instances of this type were found: *knowledge about topographical locations of process components*, *knowledge about the internal* *(technical) functioning of process components* and *knowledge about how the process is controlled.*

5. The present state of the process and the controllers. These include the present values or states of the variables and the components. It may include components which are not normally parts of the process function, for example nuts which are too worn and dirty to seat properly and so affect the efficiency or safety of the process. This kind of knowledge, although used by operators, was not explicitly looked for in the experiments reported in this study, since the present state of process and controllers could usually be read off the screen and if that was not possible, the components could be assumed to have "normal" values.

   Therefore, the present state of the process was only considered interesting if an operator observes an discrepancy between the normal and the present value of process parameters. This type of knowledge has been positioned as belonging to knowledge about *normal values of process parameters*, and is described under 2.

From the data obtained in the previous two experiments it seems as if operators are able to use different types of knowledge to answer different questions or parts of questions, and that these different types of knowledge should not be treated in a hierarchical manner. On the other hand, from the designer's point of view there are certainly relationships between structure, function and the purpose of the process (the product). Depending on the purpose of the process, certain functions are needed which are implemented by a certain structure. Since this study is mainly concerned with operator behavior, no hierarchy between the different types of knowledge is assumed.

Knowledge of the process installation can be at several levels of detail (Bainbridge, 1984). Rasmussen and Lind (1981) propose two ways of integrating process knowledge:

- **Aggregation**, which refers to the level of resolution at which parts of the plant are described, for example a pump or a cooling system. Operators work at at least two levels of aggregation. They think in terms of the individual process variables, but especially in case of incidents, they consider aggregates such as cooling or cleaning.

- **Abstraction**, which refers to the type of descriptive mode used, for example a physical component or a mass-energy flow. Rasmussen and Lind (1981) suggest that the operator is concerned with at least three levels of abstraction: the purpose of the process (why things are done), the nature of the process (what it does), and the physical properties of the process (how are functions implemented). Work by Cuney (1979) about how much operators understand about their process at a technical level showed, that less than a third of explanations given by three experienced operators were at a technical level, and that the majority of explanations were in terms of empirically observable relationships. Thus, operators are more geared towards functional knowledge about their process and the process parameters than towards the technical implementation of those functions. This is an understandable phenomenon, first of all since operators are concerned with the control of the plant in terms of process behavior. Second, once operators have identified which components are not functioning properly, further exploration and diagnosis is left to technicians. This implies that the knowledge structures of technicians with a prime responsibility for maintenance may be quite different from those of operators.

Now the different types of knowledge used by operators have been put into a framework, the question remains which kinds and how much system knowledge operators and technicians have actually available, and whether a connection can be made between the various stages of a diagnostic problem solving process and specific types of system knowledge. A number of experiments was carried out on different kinds of system knowledge to shed light on this issue. The types of system knowledge investigated have to do with the physical structure of the plant (topographical knowledge and the internal functioning of components), and the function and behavior of the process (knowledge about the process flow). In these experiments a comparison is made between expert operators, novice operators, and expert technicians. Finally, an experiment was carried out in which operators and technicians were asked to describe various process components.

7.2 Experiment 1. Topographical location of process components

7.2.1 Introduction

One of the interesting peculiarities of the paper mill studied, which was incrementally built and rebuilt, is the dissociation between the functional layout of the process and the topographical location of the various process components. This means that there is no close connection between the functional order of components and their topographical location, e.g. components that may be quite close in terms of process flow may be located in very different parts of the building, due to the incremental building and rebuilding of the paper mill and its installations.

In order to investigate the topographical knowledge available in subjects of different levels of expertise, two experiments were carried out. These experiments will subsequently be described. The prediction is that technicians will perform better than the two groups of operators, since technicians are expected to use topographical knowledge every day in their job of maintaining the installation, for which it is important to know where components are located, while the operators are more geared towards controlling the installation, which, due to centralized control rooms, does not necessarily involve knowing where the component is located.

7.2.2 Method

Subjects

Eleven subjects participated in this experiment, which is about the maximum number possible, given the limited number of experts and novices available in the area of stock preparation. They included five expert operators with more than five years of experience as independent operator, four novices who had only been two years employed by the paper mill and who had only received a general training on the various aspects of paper making, and finally two expert technicians with more than five years of experience. These technicians were included to obtain an impression about the differences in system knowledge between people with different job contents (the task of process control and troubleshooting at a production level, versus the task of maintaining the installation and performing troubleshooting activities at a technical level). The novices and one of the experts also participated in the scenario-experiment reported in chapter Five.

Materials

Components of which the location was asked are all part of the stock preparation department, but are taken from all four production lines. They consisted of the following:

- Pulper 1
- Buffer Chest K14
- Thinstock cleaners third phase
- Fiberizer 4
- Water chest C15
- Sedifloat
- Heavy pieces container
- Water chest C18
- Buffer chest K32
- Thickstock cleaners production line 3
- Contaminex
- Water Chest 3C13
- Sorter of production line 3
- Pump 16
- Buffer chest K23
- Thickstock cleaners production line 4
- Fiberizer 3F2
- Pump 13
- Reject sorter 3RS1
- Pump water chest C15.1

Procedure

All subjects were run individually in a quiet location away from the stock preparation department. They were presented the components one by one, in a fixed order and were asked to give an indication where the component was located topographically. The subjects were asked to think aloud while giving their answers. Some practice in thinking aloud and the type of question posed was given before the experiment started. The session was recorded on tape, for which the subject had given his permission in advance of the experiment. All subjects finished the session within 20 minutes.

7.2.3 Results and discussion

The tapes were transcribed verbatim and, based on answers about the various components, divided into segments. By independently consulting two shift managers, and by checking by the investigator herself, the correct answer for each component was determined. It turned out that answers could take two forms, either by giving directions how to get to a certain component, for example, "walk this way, take the stairs and then you bump right into it", or as being in a certain area, next to some other component, for example "outside, behind K31". No difference was made between those answers in the scoring procedure.

Table 7.1 gives an overview of the number of items correct for each group of subjects.

Table 7.1 *Number of components correct for each group of subjects. Results are averaged over subjects per group.*

	Experts (n = 5)	Novices (n = 4)	Technicians (n = 2)
Number correct (max = 20)	16.8	13	19.5

The Kruskal-Wallis Test Statistic T revealed a significant overall difference, T = 8.40, p = 0.01, indicating that the three groups differ with respect to the number of components correctly located. The technicians perform best, the novices worst and the expert operators perform in between. To get an indication about the difference between expert and novice operators, a Mann-Whitney U test was carried out post hoc, resulting in U = 20.00, p = 0.01, indicating that even with this small number of subjects a reliable difference between experts and novices could be established. The difference between expert operators and expert technicians resulted in U = 0.5, p = 0.07. This last result is interpreted as not significant. However, there is a trend towards a better performance of the technicians compared to the operators. It should be noted that the performance of the technicians is near perfect, and that the data show a ceiling effect. Therefore, it is very likely that there will be a significant effect if one would be able to run more subjects, which was impossible in this case, and if the data were not liable to a ceiling effect.

It can be concluded from the data of this experiment that the technicians, according to prediction, score better than the operators as a group. The novices obtained the worst score, but one would not conclude that they are unable to carry out this task: the difference in score with the expert operators is not overwhelming.

7.3 Experiment 2. Topographical location of process components in the department of stock preparation

7.3.1 Introduction

The second experiment on topographical knowledge was carried out to obtain another measure of subjects' knowledge on this matter. Therefore the subjects were asked to make a schematic drawing of all the different process components in one particular area of the paper mill: one of the floors of the department of stock preparation. This area is surrounded by walls and contains staircases to other floors, and is therefore a rather distinct area.

7.3.2 Method

Subjects

The same eleven subjects as in experiment 1 participated in this experiment.

Procedure

All subjects were run individually in a quiet location not near the stock preparation department. They were given the following question: "Please imagine that area of the stock preparation department in which the control room is located."Could you make a schematic drawing of all the process components on that floor?" Subjects were provided with an outline of the walls and one element (the control room) already drawn. They were given as much time as they needed. All of them finished within 20 minutes.

Since all subjects in this experiment had already participated in the other experiment on topographical knowledge (experiment 1), a time lag of about two months was introduced between the two experiments, to ensure as much independence as possible.

7.3.3 Results and discussion

The results were scored as follows. Each item correctly located (within certain margins) and correctly labelled (for example 1F1) was given one full point. An item correctly located but wrongly labelled (for example 2F2 instead of 1F2 which is an instance of the wrong production line, or 2F1 instead of 2F2, an instance of the right production line, but the wrong fiberizer label) was given a half point. Subjects were still given credit for this, since they know what kind of object is located in a certain place, but are mistaken about the corresponding production line, or the corresponding name.

Table 7.2 gives an overview of the number of items correct for each group of subjects.

Table 7.2 *Number of components correctly positioned for each group of subjects. Results have been averaged over subjects per group.*

	Experts (n = 5)	Novices (n = 4)	Technicians (n = 2)
Number correct (max = 26)	20.8	13.75	20.5

The Kruskal-Wallis Test Statistic T revealed a significant overall difference, $T = 6.10$, $p < 0.05$, indicating that the three groups differ with respect to the number of correctly located components. To get an indication about the difference between experts and novices, regardless of their profession, a Mann-Whitney U test was carried out post hoc with the technicians and the expert operators taken as one group and the novices being the other group, resulting in $U = 27.00$, $p = 0.01$. This implies that there is a reliable difference between the group of experts as a whole and the novices in terms of number of components correctly positioned. These results are different from the results obtained in the previous experiment in which there was also a tendency for the expert operators to score worse than the technicians, apart from the replicated finding in this experiment that novices perform worse than the other two groups. An explanation available for the difference between the results of the experiments may be that the second experiment is easier for the operators than the previous one, since the area to be drawn is the area in which the control room is located, which is thus a rather familiar area, while in the previous experiment the components were located in various places, making it more difficult for operators.

The errors made can be of three types: omissions, wrong labels, or incorrectly introduced components. This last type never occurred, but the other two occurred almost equally often. Errors in labelling are actually quite interesting, since the subject knows that a certain component is in a certain place but makes a mistake in the corresponding production line. Statements from the subject when completing the drawing indicate this process: "I know there are two cleaner batteries here, but I don't know which one goes where. It is written on the batteries themselves..". If the situation is like this in many occasions (the production line is indicated on the battery itself), then one may argue that the topographical knowledge is in fact sufficient and geared towards the task. Presumably, it is not necessary to remember the exact production line, although in many cases operators do know the correspondence between component and production line. If subjects do not know this correspondence, sometimes they are able to deduce the right answer, as the following spontaneous comment indicates: " I know that there is a fiberizer here and it cannot be the 1F2 (fiberizer 1 of production line 2) since that one is located over there, so therefore it has to be the 2F2 (fiberizer 2 of production line 2).."

Although the results presented do not indicate this point, there exist quite some differences between the drawings of the operators and the technicians. The technicians drew many electrical components (for example electrical switchboards), in fact, they started off with drawing those, since these components are considered extremely important to them, as they indicated afterwards. This can be taken as evidence for job-related knowledge about which has been reported more extensively elsewhere, in the context of expert system development (Schaafstal, 1989).

7.4 General discussion: the availability of topographical knowledge

The results of the two experiments reported above show, generally speaking, the same pattern of results. Both the expert operators and the technicians perform at a very high level. Both have good topographical knowledge about the installation. This is somewhat surprising, since it was predicted that the topographical knowledge of the expert technicians would be better than that of the expert operators, derived from the hypothesis that technicians would be more geared towards finding the exact locations of components, since that is part of their job of maintaining the installation. This tendency was only observed in the data obtained in experiment 1. In experiment 2, however, the expert operators performed at the same level as the technicians. The different results of the two experiments have been explained before. Generally speaking, the high performance of the expert operators may be ascribed to the fact that in the stock preparation department a system of centralized process control has been introduced only recently, so that the operators may still have sufficient knowledge about the topographical layout of their area. A question, which falls outside the scope of this study, is however, how long this knowledge will retain when people do not use it actively anymore.

7.5 Experiment 3. Knowledge about the process flow

7.5.1 Introduction

Another type of knowledge that was found to be used rather often while solving diagnostic problems is knowledge about the process flow. Knowledge about the process flow is related to the broader knowledge classes "definition of the product" and "process functions and behavior", discussed in the introduction to this chapter. Interestingly, it is a commonly held opinion amongst people in the paper mill that knowledge about the process flow is a very important type of system knowledge to possess, and it appears that for them this type of knowledge is the main constituent of "knowing the system". Therefore, in this paper mill, knowledge about the process flow is considered to be a quite important subject in training on the job. To investigate the expert-novice differences with regard to this type of system knowledge an experiment was carried out, in which subjects were asked to draw a part of the flowsheet of the production line they were familiar with.

7.5.2 Method

Subjects

Sixteen subjects participated in this experiment. They included two expert operators in stock preparation who were questioned about the process flow with regard to stock preparation for production line 2, four expert operators and four trainees from production line 1, three expert operators and three trainees from production line 3, three novices who have only been two years employed by the paper mill and who, in accordance with what they have been previously trained in, were questioned about the process flow with regard to stock preparation for production line 2. Finally, an expert technician participated. He was also questioned about the stock preparation of production line 2.

Procedure

All subjects were run individually in a quiet location. They were given the following question, depending on the production line they had most experience with: "Could you please make a drawing from memory of the flowsheet of your production line, starting with buffer chest 12 and ending with the headbox (for production line 1), starting with buffer chest 34 and ending with the headbox (for production line 3), or for subjects from the stock preparation department, starting with pulper 21 and ending with the pump of buffer chest 25."

Subjects were provided with paper, color pencils and a ruler by the experimenter. They were told that only the correctness of the drawing mattered, and not the neatness of the drawing. Subjects were given as much time as they needed, which varied between ten minutes (one of the experts) and two hours (one of the trainees). These differences may for a large part be attributed to the amount of effort put into the neatness of the drawing, regardless of what subjects had been told by the experimenter.

7.5.3 Results and discussion

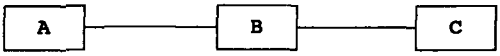
To obtain a scoring scheme for the drawings, the official flowsheets as provided by the paper mill were divided into several components: chests and other big components (e.g.

fiberizers, reject sorters), pumps, controls, pipings between components, and "bypasses", which are pipings between components from different production lines, for example a piping from the pump of chest 25 of production line 2 into buffer chest K13. These bypasses were included since they are rather important sources of knowledge for troubleshooting. Sometimes bypasses offer possibilities for keeping the process running when a component is not working properly, by means of sending flow along a different route than it normally would follow. This scoring scheme resulted in the following number of elements:

Table 7.3 *Number of elements in the flowsheets of all three production lines.*

	Comp.	Pumps	Contr.	Pipings	Byp.	Total
Prod.line 1	16	5	5	14	8	38
Prod.line 2	15	6	3	14	2	40
Prod.line 3	16	5	7	19	--	47

The drawings were scored as follows. Firstly, a score of one was assigned to each element of the above categories if they could be identified in the drawing. Pipings were the hardest cases, since they were identified as pipings between two elements. If one of the elements was lacking in the drawing, the following solution was chosen. Imagine the following situation:



In this situation, A, B, and C are components and the lines in between them are pipings. Thus, this drawing consists of the following elements: A, B, C, A-->B, B-->C. If A and C are drawn, but B is not, and a subject draws a piping between A and C, he is assigned a score of 3, two for the components and one for the piping.

To enable comparisons between various production lines, the scores per subject were calculated in percentages of maximum number of points correct. These results have been averaged over groups of subjects and are shown in the following table.

Table 7.4 *Number of elements in flowsheets, expressed in percentages, correctly positioned by different groups of subjects, for different production lines.*

	Comp.	Pumps	Contr.	Pipings	Byp.	Total
Prod.line 1 experts	98.3	100	55	96.5	59.5	83.5
Prod.line 1 trainees	98.3	100	50	93	12.5	73.0
Prod.line 2 experts	100	100	50	86	75	90
Prod.line 2 novices	100	100	45	93	0	86.7
Prod.line 2 technician	40	0	0	7	0	17.4
Prod.line 3 experts	98	80	57	89.3	--	83.7
Prod.line 3 trainees	92	77.7	19.3	84	--	74.3

One of the interesting results of this experiment is the difference between the operators on the one hand and the technician on the other. The technician was not able to complete his drawing due to a lack of knowledge: he gave up. This is even more interesting since he himself believed to be able to do this task. We tried this experiment with him on all production lines (the table presents only the results on production line 2 with which he was most familiar), but he was unable to produce a reasonable drawing on any of them. For further analyses his data are discarded.

A Chi-square test revealed no significant differences between the totals of experts and trainees on any production line. However, these are overall results (percentage of elements correct), and it is questionable whether it is justified to sum up all elements without taking into account the differences between the elements. Therefore, separate analyses were carried out for each type of element (components, pumps, pipings, controls and bypasses) over production lines. The reader should notice that differences between experts and novices with regard to bypasses could only be based on production lines 1 and 2, since in the relevant part of production line 3 no bypasses occur. These results revealed no significant differences for components, pumps, controls, and pipings, indicating that both experts, trainees, and novices score equally well on those subjects. The only significant result that was found was on the bypasses. A Kruskal-Wallis analysis of variance revealed a significant effect, KW = 8.56, p = 0.01, indicating that on this element, the experts perform significantly better than the trainees and the novices. In fact, only one of the trainees mentioned some bypasses (50%), but none of the drawings of the other trainees or novices showed any sign of this type of element. This is a rather interesting result since it would seem to suggest that bypasses are a sign of an integrated representation of the production lines which has not been built up yet by trainees and novices. The fact that all groups score equally well on the other elements is also striking. However, the fact that both novices and trainees perform as well as experts may partly be attributed to ceiling effects. On a number of occasions, perfect scores are reached which may diminish existing differences. On the other hand, the non-difference may be attributed to the fact that both the trainees and novices are still in a training situation and familiarizing oneself with the flowsheet is an important part of this training. However, the very good performance of both the trainees and novices on this important type of system knowledge shows no relationship to their very impoverished performance when it comes down to effectively using this knowledge in troubleshooting situations as reported in chapter Five and Six.

If the scores on the different element groups are compared between production line 1 and production line 3 it appears that there is a trend towards people from production line 1 scoring higher. However, this trend does not reach significance. It would have been interesting, though, since one of the differences between both production lines consists of the way the operator controls the installation: on production line 3, the operator has the flowsheet in front of him on a computer screen and he is able to control elements directly from there, while on production line 1 the operator has much more direct contact with the process, and should do a lot more walking around to control several elements. This situation shows some comparisons with experiments carried out about differences in knowledge acquired from maps and by direct navigation through the environment. These experiments show that with extensive experience knowledge acquired through navigation is superior than knowledge acquired from a map (Thorndyke and Hayes-Roth, 1982). Another reason why this trend is interesting is that it shows that, although operators have the flow sheet in front of them all day, they are not able to perform better when asked to draw it from memory than their colleagues who cannot use this aid when building up this representation.

7.6 Experiment 4. Knowledge about process components

7.6.1 Introduction

The final type of system knowledge that was investigated was knowledge about the structure and function of process components which belongs to the broader class of "knowledge about the plant's physical structure", in particular knowledge about how functions are obtained. In general, it is known that people can have a very rich set of kinds of knowledge about a particular piece of equipment. Table 5, based on studies reported in Kieras (1982) lists some of these different kinds of knowledge. This table lists some of the types of knowledge that can be identified when subjects, both experts and novices, were asked to describe actual pieces of equipment, ranging from devices as simple and familiar as alarm clocks, to complicated and unfamiliar pieces of electronic laboratory equipment.

Table 7.5 *Types of knowledge people have about devices (adapted from Kieras, 1982).*

The label or name of the device
The functions or purpose (what goals can be accomplished)
The controls and indicators
The inputs, outputs, and connections
Power source and requirements
External layout and appearance
Internal layout and appearance
External behavior (input-output function)
How to operate the device to accomplish goals
Procedures for troubleshooting and maintenance
Internal structure and mechanisms (how it works)

As suggested by table 7.5, most knowledge about devices seems to be related to *using* the device, as opposed to *how-it-works knowledge* about the internal structure and operation of the device. For example, Kieras (1982) observed that when asked to freely describe a device, experts would provide considerable detail on the procedure for using a piece of equipment, but often they did not consider it necessary to provide any details about how the device worked, although it was clear that they would be able to do so when asked.

The question is whether these results can be replicated with process components taken from an industrial environment and whether the same kinds of knowledge are used in this situation. Another question is whether people with a different profession establish different kinds of knowledge about process components. The protocol studies reported in chapters Five and Six indicated that people use knowledge about the functioning of process components in diagnosis, which makes a third reason for carrying out an experiment on the kinds of knowledge people have about process components. Finally, this experiment enables us to investigate what types of knowledge are available when someone does not have much experience with the component. This situation is created by asking an expert paper maker to describe components from the stock preparation department, and asking an expert in stock preparation to describe components belonging to the paper machine.

7.6.2 Method

Subjects

Six subjects participated in this experiment. They included three novices who have only been employed for two years by the paper mill and have only received a general training on all aspects of paper making, two expert operators with more than five years as independent operator. One of them is an expert in stock preparation, one is an expert paper maker. Finally, one expert technician participated.

Materials

Subjects were asked questions about the following process components:

- Fiberizer
- Reject sorter
- Reject drum
- Vertical screen
- Refiner

The vertical screen and the refiner are part of the responsibility of the paper maker, the other three, fiberizer, reject sorter and reject drum, belong to the domain of stock preparation.

Procedure

All subjects were run individually in a quiet location. They were given the following questions: "Could you please freely describe for me the following component", followed by the first component. The next question was: "Please describe the internal working of this process component". This question was included to get some indication about the subjects knowledge about the internal workings, and is a test of Kieras (1982) assumption that operators are able to describe the internal functions. The final question posed to the subject was "Please describe as many causes for malfunctioning of this component as possible".

The novices and the technician all were questioned about the components in the same order. The expert paper maker started with the components that belong to his department, followed by the other three, while the expert in stock preparation started off with the components of stock preparation, followed by the two components belonging to the paper machine. Thus, the expert paper maker had to describe three components that he is not very familiar with, and the expert in stock preparation described two unfamiliar components. If subjects wanted to, they were allowed to make drawings of the components, for which paper and pencils were provided. Subjects were given ample time for their descriptions. They all finished within 30 minutes. The session was recorded on tape, for which the subjects had given their permission prior to the experiment.

7.6.3 Results and discussion

The data obtained are analyzed for each process component and subject separately in the following way. The overview of types of device knowledge as derived from Kieras (1982) is used as a framework to interpret the results. Some changes to this framework have been made, to further the applicability in the current situation. "Input/output and their connection" is interpreted as meaning the input and output for the component. "Procedures for troubleshooting and maintenance", an indication of troubleshooting knowledge, has been changed into number of malfunctions. "Power source and requirements" is interpreted as requirements for the components for good operation. This may apply to power source requirements, but also to process requirements, such as a stable input pressure. Table 7.6 gives an overview of the results obtained.

Table 7.6 *Knowledge types used by different subject in describing process components. Exp. 1 is the expert in stock preparation, while exp. 2 is the expert paper maker.*

	Nov.1	Nov.2	Nov.3	Exp. 1	Exp. 2	Techn.
Name	6/6	6/6	6/6	6/6	6/6	3/3
Purpose	6/6	5/6	6/6	2/6	6/6	0
Controls	4/6	3/6	0/6	2/6	1/6	2/3
Input/Output	6/6	6/6	6/6	5/6	3/6	3/3
Power source/ requirements	0	2/6	0/6	2/6	4/6	3/3
External appearance	6/6	6/6	6/6	6/6	6/6	0
Internal appearance	5/6	6/6	6/6	4/6	4/6	3/3
External beh.	6/6	6/6	6/6	4/6	4/6	0
Operation	6/6	2/6	4/6	4/6	4/6	0
How it works	6/6	6/6	6/6	4/6	4/6	3/3
No. of malf. (mean)	2.8	1.8	3.7	2.2	3.5	5.3

Due to time limitations descriptions of only three components could be obtained from the technician. The results show that Kieras' (1982) results could be replicated fairly well, since there is no knowledge type that has never been used, and all replies could be accomodated. However, one should bear in mind that the framework has been slightly adapted beforehand. The results also show that, in line with Kieras' statement, subjects are able to provide how-it-works knowledge when asked so. In Appendix B, examples are given of the qualitative data obtained on the process components. It is interesting to note that in this situation internal appearance is coupled with knowledge about how it works. Subjects make a connection between various internal subcomponents and their function. For example: "the sorting cylinder contains high pressure sprinklers, with the following function: cleaning and thinning" (expert 1). Sometimes a connection is made between the external appearance and the function, for example "the holes in the wire of the fiberizer are cone-shaped to ensure a differential pressure". In a general sense, subjects give descriptions of the components in terms of the interaction between structure and function. No stable differences between expert operators and novices could be found, presumably since the novices had recently been learning these components for one of their exams. In fact, the novices gave fairly extensive accounts about the various components and, compared to the diagnostic tasks, did not find this task a very difficult one.

What types of knowledge are available when someone lacks concrete experience with the component? It can be concluded from the data as presented in table 6 (non-perfect scores of the expert paper maker and expert in stock preparation on a number of knowledge types, based on imperfect knowledge about unfamiliar components) together with appendix B, which contains examples of answers given by subjects, that at least name, function and external appearance are still available, but that the other types of knowledge will not emerge. One may conclude that for an operator especially name and function of a component are basic knowledge types. The external appearance of the component is a relatively simple one, since everybody sees every component every once in a while. For an operator, one would have expected a higher emphasis on operation of the component. The reason that this does not show up may be that in this situation operation of the components merely consists of turning on/off whilst other types of operation have to do with rather subtle changes in settings that are not needed in all circumstances.

Finally, the descriptions of the technician are rather different than the descriptions given by the group of operators. The technician provided his descriptions at a technical level, and included much more technical details such as valves and their accompanying controls than any of the operators. Also, the type of malfunctions that he mentioned were rather different from the ones mentioned by the operators, such as bad pulleys, malfunctioning of control valve, and leaking flange, while the operators mention malfunctions such as inlet clogged, sprinkler clogged, and inlet pressure too high.

7.7 General discussion: the availability of system knowledge

In this chapter, experiments were reported about the availability of different types of system knowledge in different groups of subjects. The types of system knowledge investigated were topographical knowledge, knowledge about the process flow, and knowledge about process components. From the experiments reported a rather diverse pattern of results emerged. Experiment One and Two showed that both expert operators and technicians perform at a high level with respect to topographical knowledge. The novices do less well, but are still able to obtain a reasonable score. The experiment concerned with knowledge about the process flow showed that the most important difference between operators with varying levels of expertise was the difference in integration as measured by the number of bypasses. Apart from that, no differences could be found between expert operators, trainees and novices. It turned out that the technician had to give up on this task: he could not do it. The final experiment in which people were asked to describe various process components, showed no systematic differences between expert and novice operators, but again an interesting difference between operators and the technician, which is in line with a study reported before (Schaafstal, 1989). From these results as a whole, it is fair to conclude that there are qualitative differences in the way people with different job-contents have represented their knowledge. Technicians work at a different level of aggregation (Rasmussen and Lind, 1981) than operators. They are much more concerned with technical implementation of process functions and technical details in general than operators, which is consistent with the results obtained by Cuney (1979). This difference in representation of the installation between operators and technicians has implications for several areas, such as the development of knowledge-based system and training issues, which will be discussed in the next chapter. Looking at differences between expert and novice operators it may be concluded that, although sometimes present, they are not overwhelming, and thus, differences in availability of system knowledge cannot satisfactorily explain the differences between experts and novices in troubleshooting.

Chapter 8

General discussion

8.1 Introduction

The preceding chapters aimed at giving an answer to the following questions: a) what is diagnostic skill in process operation, b) which types of system knowledge play a role in diagnosis in paper mills and c) how does diagnostic skill develop or, alternatively, what differences can be found between expert and novice operators with respect to diagnostic skill?

The notion underlying this study is that in general, skilled behavior may be attributed to well-structured declarative domain knowledge, coupled with efficient problem-solving strategies. Applying this framework to diagnostic skill resulted in the development of a model for a task-level diagnostic strategy as described in chapter Three. The model consists of a description of the different steps taken in diagnosis, coupled with a specification of the ordering of those steps. The model aims at giving an account of human behavior with respect to diagnosis, as opposed to an idealized account of problem solving behavior that could be used as specification for a knowledge-based system. This model has been tested in experiments using verbal protocols reported in chapter Five and Six. It was found that the steps predicted in the model were very much in accordance with the steps that could be deduced from the verbal protocols obtained from expert operators. However, it turned out that the diagnostic strategy applied by novices as well as trainees was rather different from the one employed by expert operators. Both novices and trainees do not systematically make judgments about the seriousness of the situation, and they also "forget" to evaluate their problem solutions. In general, both expert and novice operators do behave according to the predicted ordering of steps. However, the behavior of both expert and novice operators is more in accordance with a depth-first strategy with respect to ruling out possible faults as likely candidates for the problem solution than expected beforehand.

Evidence could be obtained for a, maybe partial, dissociation of the development of a strategy for diagnosis and the availability of domain-specific system knowledge. First, transfer subjects, that is, expert operators questioned about a rather unfamiliar area of paper making, were able to employ a diagnostic strategy very similar to that of expert subjects in that area. Second, the experiments reported in chapter Four and Seven showed that inexperienced operators do not necessarily lack system knowledge. Sometimes they even perform at the same level as their expert colleagues with respect to tasks concerning domain knowledge. However, when they have to solve diagnostic problems (chapter Five and Six), this knowledge is not brought to bear, or is insufficient in helping them to solve problems in a systematic way. Thus, although system knowledge may be necessary to support a diagnostic process, it certainly is not sufficient to enable the use of a good diagnostic strategy. Various types of system knowledge could be identified that are used in diagnosis as reported in chapter Five and Six, such as knowledge about the process flow, topographical knowledge, knowledge about paper making, its control, and the process dynamics. The experiments reported in chapter Seven give evidence for the statement that system knowledge is partially job-dependent: differences were found between the system knowledge of technicians and operators.

In what follows, the applicability of the diagnostic model for the development of expert systems and operator training issues in paper making will be discussed, followed by implications of the research reported in this study for the development of knowledge-based systems and operator training issues in a more general sense. Finally, the implications of this research for theories about the development of cognitive skill will be discussed.

8.2 Applicability of the diagnostic model

The model for diagnosis as initially described in chapter Three and tested in experiments reported in chapter Five and Six was defined at the task level, providing an account of the structure of a diagnostic process. As such, the model may be useful as an interpretation model in the knowledge acquisition process for the development of knowledge-based systems. However, one should keep in mind that the model has only been tested in one domain, paper making, and to obtain a better idea about the full scope of its generality, it should be tested with data from other domains as well. In this respect, it is interesting to compare the model to the model of reasoning developed by

Wognum (1990), since that one has been tested in the medical domain. The tasks that Wognum distinguishes in diagnosis are: Develop differential, Generate causes, Consider hypothesis, Discriminate hypotheses, Check findings, Refine hypothesis, and Advise diagnostic test. These tasks are comparable to several of the steps taken in diagnosis as defined in chapter Three. "Generate causes" in Wognum's model is similar to "Possible faults", "Discriminate hypotheses" is rather similar to "Ordering of Faults according to likelihood", and "Check findings" is similar to "Testing". One of the reasons that the models are not completely comparable is that Wognum concentrates on the process from identification of symptoms to the development of the diagnostic differential, while our model aims at describing the reasoning process up to an evaluation of the solution. Thus, it has a wider scope than Wognum's model. Another difference between the two models is the emphasis put on psychological validity. Wognum's model was not developed explicitly as a framework to account for human behavior, it is a framework for developing competence models. Our model, by defining all legitimate transitions, certainly emphasizes the description and explanation of human behavior, it is a performance model. Although competence models may suffice as interpretation models for knowledge structuring processes, psychologically valid models are needed to make fair statements about human behavior. This will be especially important in those situations requiring advice about the capabilities of human beings, such as may be found in man-machine interface issues, and in operator training programs.

The test that Wognum performs on her model by fitting it onto medical texts, in which an idealized form of problem solving is presented, shows that her model gives a fairly good account of the contents of the medical texts. Since Wognum's model is applicable in the medical domain, it is assumed that, due to the similarities to Wognum's model, our model has a wider applicability than paper making or process control in general.

It may be, though, that in certain domains a different emphasis is put on the different steps. For example, the identification of symptoms may sometimes be rather hard, but was often an almost trivial matter in the diagnostic tasks reported in this study, since it always involved alarm handling situations. Since many faults underlying symptoms in the paper industry can be described as "filthy or broken <something>", the appropriate repair is not hard to find: "clean or replace <something>". On the other hand, the enumeration of all possible faults is a rather complex task in the paper industry, and this may be easier in some other domains. Thus, it would certainly be recommended that when applying the model in a new domain to tune the model towards the specificities of that domain.

The model is geared towards only one aspect of diagnosis: a description of the structure of diagnosis at the task level. To obtain a more complete model of diagnosis it should be integrated with search strategies which describe how in particular situations operators make connections between symptoms and underlying faults, or more in general, how transitions between the various steps of the model are accomplished. Different search strategies have been described in the literature, some of which may be considered more powerful strategies than others. Obviously, heuristics (or symptomatic search), connecting symptoms to underlying faults, belong to the most powerful strategies an operator may have, although they are only applicable in a narrow domain, and have no wider generality than this domain. Therefore, they are likely to fail in any new situation. Less powerful, but still leading to conclusions rather efficiently, are search strategies such as topographic search, geared towards diagnosis in technical domains, but more widely applicable than heuristic search. At the next level of generality are search strategies such as split-half approaches, in which the goal is to minimize the number of tests to localize the source of the failure. Experiments by Goldbeck, Bernstein, Hillix, and Marx (1957) showed that training subjects to make use of a split-half approach only enhanced performance in simple systems. This is understandable since a split-half approach, due to its generality, lacks the power to handle complex systems efficiently. In simple systems, though, one may not need such powerful strategies to be able to come up with solutions efficiently. At the lowest level of specificity are so-called weak problem solving methods, such as means-ends analysis and generate and test, as described by Newell and Simon (1972). These search mechanisms are asserted to be a central component of general problem-solving skill and are very general in scope, thus trading power for generality.

The search strategies used by operators may be different in different situations, and may also depend on the level of expertise. Since heuristics are mostly developed through practical experience and are tied to specific situations, they may become increasingly available with increasing levels of expertise. As demonstrated by Rasmussen (1976), topographic search strategies, although not informationally economic, may be preferred by technicians in domains such as electronic troubleshooting. In the experiments reported in chapter Five and Six, heuristics are used rather often by the experts, but there are also many instances of functional search, in which, based on the functioning of the installation series of good/bad judgments are carried out on individual process components. Novices mostly rely on heuristic rules, insofar as available, but they also rely on functional search, and now and then even on weak methods such as means-ends analysis. Thus, for every specific diagnostic situation, the different search strategies employed should be investigated.

Apart from the search strategies that are used by different subjects, the model should be completed with a specification of underlying system knowledge for each step of the model. Which types of system knowledge are used in each part of the diagnostic process? Chapter Six gave us some clues regarding this issue.

Since the types of knowledge identified are obtained in experiments with operators as subjects, one should keep in mind that an operator-oriented point of view was presented, which may be rather different from a technicians' point of view. The types of knowledge used by technicians presumably are different than those used by operators. Chapter Seven provided us with insight regarding those differences, since the technician in these experiments was not able to perform reasonably on tasks concerning knowledge about the process flow. On this basis, one may deduce that he does not have to use this knowledge very often. On the other hand, technicians are expected to have a much better knowledge about the functioning of process components, since that is what they are mainly dealing with.

In conclusion: the task-level model for diagnosis as presented is assumed to have wider generality than just the domain of paper making. However, only one aspect of diagnosis, the task-level structure, is discussed, and to obtain a full model for diagnosis, information about how transitions between the various steps are accomplished should be added, coupled with an investigation about the relation between the diagnostic steps and the underlying system knowledge.

8.3 Implications for the development of knowledge-based systems

As stated before, the model as such may be used as an interpretation model for knowledge acquisition and knowledge structuring during the development of knowledge-based systems for diagnostic tasks, in the same vein as the KADS interpretation models (Breuker et al. 1986), discussed in chapter Three. The main advantage of the model presented here over the KADS interpretation models is its relation to real-life situations, which makes it easier to apply, since no translation from a rather abstract model to practice is needed.

Many of the knowledge-based systems built nowadays only rely on heuristic knowledge. Although these systems may be very powerful, they are not very flexible, because their power is restricted to a very narrow domain and their robustness is low. Recently, interest has grown in diagnostic systems that utilize deep knowledge of the system under diagnosis. These systems use an explicit model of the device and can predict, verify or explain behavior of the device from first principles (Davis, 1984). In so called *second generation expert systems* a combined use is made of shallow knowledge (heuristics) and a deeper understanding of the device. Although the use of deep knowledge makes diagnosis less efficient, it greatly enhances robustness and more over supports deeper explanation of the diagnostic process. However, the question remains how deep and shallow knowledge should be combined in such systems. The experiments reported in this study provided us with some insight into this question, since evidence could be obtained about the types of deep system knowledge that were used in realistic diagnostic situations. It turned out that operators are very flexible in the type of knowledge they use in a particular situation. Thus, they employ several models of the installation at different levels of abstraction, presumably dependent on the difficulty of the problem: a more difficult problem requires a more detailed model of the device. Also, the type of system knowledge used presumably depends on the phase of the diagnostic process: symptom identification may require different types of system knowledge than the identification of possible faults, as suggested in chapter Six.

The two-level reasoning model as developed by Wognum (1990) provides us with a nice framework for describing the issues regarding the cooperation between heuristic and deep knowledge. In this two-level reasoning model, a difference is made between reasoning at the level of the global task structure and reasoning at domain level. The task level serves as a control structure for the reasoning process. The model developed in this study, based on the same idea, gives a fuller account of diagnostic reasoning at the task level. Reasoning processes at the domain level consist of local strategies for implementing the goals from the global task structure, coupled with domain knowledge. Reasoning at task level and at domain level is an alternating process. A problem-solving process starts with reasoning at task level, which in turn calls for a reasoning process at domain level. When a goal has been reached the result serves as an input for the next subtask, and so on. The search strategies as identified by Rasmussen (1976) and others (e.g., Benjamins & Jansweijer, 1989) may be viewed as local strategies used at domain level. Since humans are flexible in the application of strategies, it may be considered useful to implement various search strategies in second generation expert systems as well. Moreover, the model described in this study may serve as a basis for the investigation of the local search strategies used at each point of the task level employed in various diagnostic domains. The models of the installation used by operators, which are mainly functional in nature, may provide us with hints regarding the use of domain knowledge at various stages of the diagnostic process.

8.4 Implications for operator training issues

An interesting issue with respect to the model is the question whether it can be used for operator training. Before this point will be addressed, it is worth mentioning some comments about the present way of training operators. An important feature of current operator training courses, both on-the-job and off-the-job courses, is the heavy emphasis on the acquisition of system knowledge. However, this is too often approached using theoretical principles or formal scientific laws that define the domain. Results of several studies, as summarized in Morris and Rouse (1985) indicate that instruction in theoretical principles is not an effective way to produce good troubleshooters. Explicit training in theories, fundamentals or principles fails to enhance and sometimes actually degrades diagnostic performance (Brigham & Laios, 1975; Crossman & Cooke, 1974; Kragt & Landeweerd, 1974; Morris & Rouse, 1985; Mayer & Greeno, 1972; Mayer, Stiehl, & Greeno, 1975). However, one cannot make a full diagnosis without having system knowledge, but the types of system knowledge trained should be geared towards the job one has to carry out. For the paper industry, the types of system knowledge used by the expert operators may serve as a starting point for the identification of relevant pieces of knowledge that are used in practical situations, and which are therefore certainly worth training. Despite the importance of good diagnostic skills in process control, not much effort is put into explicit training of good diagnostic strategies. This may partly be due to the misconception that training in system knowledge will automatically result in good troubleshooting performance, since the two are closely linked. It may also stem from the idea that good troubleshooting skills will automatically evolve with increasing experience on-the-job, and therefore explicit training will not help all that much. In itself, this idea is valid: increasing experience with troubleshooting most of the times leads to a better performance: the expert operators in this study became fine diagnosticians without explicit strategy training. However, the question is whether operator training as a whole can be speeded up and be made more efficient if strategy training is taken into account. A final reason for the absence of strategy training is the fact that good strategy training appears difficult to accomplish, which makes it hard to incorporate strategy training in regular operator training courses.

Now having discussed some of the issues in current operator training, we will continue with a discussion about the topic whether the model presented can be used for strategy training. Since the model has been defined at a task level, it basically consists of a goal structure, in which every phase or task as defined can be regarded a goal. Thus, the model offers a guideline and opportunities for training explicit goal structures. This is especially interesting since, although importance of strategic control knowledge in the form of goal structures has been shown in a number of domains (Greeno, 1978; Card, Moran, & Newell, 1983; Anderson, Farrell, & Sauers, 1984; Kieras, 1987; Gott & Pokorny, 1987), it happens that in instructional materials a number of times gaps have been identified with respect to goal structures (Greeno, 1978; Anderson et al., 1984; Kieras, 1987). The emphasis on the importance of training people how to use strategic knowledge is also one of the most important outcomes of the Guidon project (Clancey, 1987). Training of explicit goal structures is not impossible, though, since it has already been accomplished a number of times, for example in some of the intelligent tutoring systems, such as Proust, a knowledge-based cognitive diagnosis program that identifies programming bugs with an emphasis on bugs regarding the abstract program plan (Soloway & Johnson, 1984), Bridge, a tutorial learning environment that concentrates on the planning knowledge required early in programming skill acquisition (Bonar, 1985), and the Lisp tutor (Anderson & Reiser, 1985). For Proust, it has been shown that its effectiveness in diagnosing programming errors on simple Pascal problems is comparable to that of human teaching assistants, that is, about 75% correct. The Lisp tutor has been evaluated involving comparisons of various instructional alternatives to learning Lisp, including traditional classroom instruction, a private human tutor and independent self-study. In these studies, the Lisp tutor is not far behind human tutors in effectiveness. The self-study alternative was much worse, particularly as the instructional material became more difficult. When a group of students used the Lisp tutor for problem exercises to supplement standard lectures, they spent 30% less time on the problems than students working on their own, but scored 43% better on the post-test (Anderson, Boyle, & Reiser, 1985). Pirolli and Anderson (1985) examined the particular effectiveness of teaching a Lisp goal structure as a strategy for learning recursive functions as compared to "process" instructions that explained how recursive functions work, but did not teach specific steps in writing them. The "structure" group was 32% faster than the "process" group in correctly generating an initial set of functions. Anderson (1987) argues that instruction in a skill is most effective when it directly provides information needed in a production-system model of that skill. This information includes the goal structure as well as procedural steps. The results regarding these computer tutors may be considered promising with respect to strategy training.

With respect to training troubleshooting procedures, quite some research has been carried out on the effect of *proceduralized training*. Proceduralized training represents an extreme form of rule-based troubleshooting instruction. The student is provided with exact step-by-step procedures at a very concrete level to learn and practice during training sessions. The theory of performance that underlies this approach is problem solving as "the rote execution of procedural steps". Learning is assumed to occur via rote practice and general physical familiarity with the system. Potter and Thomas (1976), in a study regarding troubleshooting performance in a technical domain, showed that fully proceduralized job aids resulted in superior performance compared to technical orders or logical troubleshooting aids. This effect was strongest for simpler fault isolation tasks and for less experienced technicians. This suggests the same kind of instructional applicability as seen in the Lisp tutor (Pirolli & Anderson, 1985) where a complete prespecification of steps appears feasible and effective on easier tasks in the initial stages of learning, but may loose effectiveness for more difficult tasks. This result may be explained by the absence of an explicit goal structure in proceduralized training, while an explicit hierarchically organized goal structure is one of the characteristics of highly efficient expert behavior, which will especially be called upon in more complex situations. A pragmatic limitation to the proceduralization approach is the fact that one can seldom anticipate all events or combinations of events that may occur in a particular system. Thus, operators will inevitably encounter events for which there is no such procedure, or events in which it is unclear which procedure, if any, should be used. In such situations, proceduralized training is of no use (Rouse, 1982). Morris and Rouse (1985) therefore conclude their review of research in troubleshooting with the recommendation that the most promising instructional approach worthy of further research is one where procedures for approaching fault isolation problems are combined with instruction in how to use system knowledge in deciding exactly what to do.

Translating this to the situation in the paper industry, and process industry in general, leaves us with the following recommendations for operator training. First, system knowledge to be trained should be geared towards job contents, and thus, a thorough analysis of job contents is of crucial importance. This implies that the system knowledge taught to technicians and operators may be rather different in nature. For operators, it is especially important to know aspects such as the function that certain process components have on the paper making process, but they should not be bothered with too many technical details. Technicians on the other hand, need a thorough understanding of components at a technical level, and their knowledge about the influence of components on paper making may be more superficial. Second, much more effort should be put into training of an explicit task-level diagnostic strategy, since such an efficient goal structure appears to be almost completely lacking in novice behavior. This will enhance training efficiency, job efficiency, and thus production efficiency.

8.5 Models for diagnosis: widening the scope

As stated before, the model developed in this study is aimed at only one level of diagnosis: the task structure, and thus, the model is of a somewhat limited scope. To obtain a full scope model of diagnosis, various other elements have to be added. This section will discuss some of these elements. The task-level model consists of several steps, which may be taken as high-level goals in a diagnostic task. Goals are fulfilled by means of local search strategies (subgoals) such as described by Rasmussen (1976). If necessary, relevant domain knowledge has to be applied to obtain values for those subgoals. This view of diagnosis as a somewhat hierarchical process opens a number of research questions. First, the question is which search strategies are applied, and are effective in each part of the diagnostic task. To determine which repair to take in a given situation may be a rather different process than fault finding, and thus, presumably, different local strategies may play a role in the two tasks. The work by Rasmussen and Rouse, as discussed in section 3.3 bears relevance to this question. The second question to ask would be the interaction between local search strategies and types of underlying domain knowledge used. Presumably, not all kinds of domain knowledge play a role in each diagnostic step, and thus, one would like a finer-grained experimental analysis of used domain-knowledge (system knowledge) in different diagnostic steps than was possible in this study. *Chapter Six* contains data of an exploratory nature that are relevant in this aspect. Research by Kieras among others, described in chapter Seven, is relevant as well. Finally, not much is known yet about the generality of certain aspects of system knowledge. Although a lot of research is carried out about structural and functional models, the question is whether these are the only two possibilities, or whether different models, for example causal models, may be used as well. A related question concerns the flexibility employed by people in switching between various models. The protocol studies reported in chapter Five and Six showed that people use, depending on the situation, different types of system knowledge. Thus, in modelling human diagnostic behavior this flexibility in using system knowledge should be taken into account as well. The study presented was not aimed at giving an answer to all these questions, and restricted itself to a description of the task level of diagnosis. However, it would be interesting to search for coherent answers in the future to some of questions posed above. For the moment, the hierarchical view of diagnosis presented in this section may be used as a framework for task analysis, on the basis of which decisions can be made upon where and how to support operators and which makes that an analytic approach to operator training becomes possible. Besides this, it may be used in general for knowledge structuring process.

8.6 The development of diagnostic skill

From the research reported in this study it appears that diagnostic skill comprises at least three components: a global task structure geared towards diagnosis, appropriate local search strategies for the implementation of goals set in the global task structure, and domain knowledge to implement the local search strategies. The view emerges that the domain knowledge, the models of paper making and the installations used in paper making, is acquired more or less before and independent of the development of the control structure. These findings will be explained in terms of the ACT* theory of skill acquisition (Anderson, 1982, 1983, 1987; Singley & Anderson, 1989) later on. Before this is done, first a brief overview of the most important features of this theory will be given. For a more detailed overview, the reader is referred to Anderson (1987) and Anderson and Singley (1989).

The ACT* theory of skill acquisition has been implemented as a production system model, in which a distinction is made between declarative knowledge (knowing that) and procedural knowledge (knowing how). In the course of skill acquisition knowledge comes in a declarative form and is used by weak methods to generate solutions, and, based on these solutions, the knowledge compilation process forms new productions. Thus, knowledge will gradually be proceduralized, resulting in efficient problem solving procedures. Important features of the ACT* theory are the following:

- *Productions form the units of knowledge.* Productions define the steps in which a problem is solved and are the units in which knowledge is acquired.
- *Hierarchical goal structure.* The ACT* production system specifies a hierarchical goal structure that organizes the problem solving. The hierarchical goal structure closely reflects the hierarchical structure of the problem. Goals are also important in structuring the learning by knowledge compilation. They serve to indicate which parts of the problem solution belong together and can be compiled into a new production.
- *Initial use of weak methods.* The use of weak methods, such as analogy or means-ends analysis, plays a critical role in getting initial performance off the ground.
- *Knowledge compilation.* All knowledge in the ACT* theory starts out in declarative form and must be converted into a procedural format. This declarative knowledge can be encodings of examples of instructions, encodings of general properties of objects, and so on.

Based on the principles in this theory, Anderson and coworkers built several computer tutors for fairly complex domains, such as programming Lisp, solving geometry proofs, and solving calculus problems.

If the ACT* theory of skill acquisition is applied to the findings reported in this study, the following conclusions can be drawn. Initially new operators start off with the acquisition of pieces of declarative domain knowledge, which will gradually be assembled into more coherent units. As they become involved in troubleshooting activities, initially they will just be involved in the undertaking of simple actions, which will result in the development of procedures for simple tasks (for example procedures for changing pressure, or procedures for recognizing certain symptoms). However, these procedures are still tied to very limited domains and novice operators may not know the conditions of effective application of this knowledge since they have not been told so. The major question is how the hierarchical control structure develops. According to Anderson's theory, this is just a side-effect of the learning process since the control structure strongly reflects the task characteristics. With increasing experience, problem solving activities will become more and more adapted to the particular task, and thus, the control structure will appear as side-effect of experience in problem solving. Other researchers (for example Simon & Simon, 1978; Larkin, McDermott, Simon, & Simon, 1980; Miyake & Norman, 1979) put much more emphasis on the development of this control strategy as a form of self-regulatory skill (summarized in Glaser & Bassok, 1989). In this line of research the question is asked whether general cognitive skills exist, which may be applied to a wide range of problems and which can, in fact, be trained. Examples of these general problem solving skills can be found in courses on improving general thinking skills as reported in Hayes (1985). However, the differences between this approach and the approach taken by Anderson and coworkers may not be as great as it may appear, since the computer tutors built by Anderson and coworkers contain explicit goal structures and solution strategies to keep students on the right track. Presumably, the most important difference between the two approaches is in the supposed level of generality of problem solving skill. Anderson and coworkers are not very optimistic with respect to transfer possibilities of general cognitive skills (Singley & Anderson, 1989), but other researchers have more faith in this approach (for example Simon, 1980; Papert, 1980; Brown, Bransford, Ferrara, & Campione, 1983). However, it has been very hard to demonstrate transfer of general problem-solving skills across a wide range of problems (summarized in Singley & Anderson, 1989). The problem underlying this finding may be that general methods are often useless in problem solving because their prescriptions, although intuitively appealing, are too vague to apply and the student is unable to relate them to the problem at hand (for example prescriptions like "use analogies", or "when you're stuck on a problem it is wise to take a break and do something else".). However, problem solving methods at a moderately general level have been discovered and successfully taught in the areas of mathematical problem solving, deductive logic, statistical reasoning, and computer programming (for a review of these studies see Singley & Anderson, 1989). Thus, as long as the correct level of generality is chosen, which may be deduced from a task analysis such as the one carried out by Klahr and Carver (1988), transfer of moderately general problem solving strategies to other domains is possible.

The hierarchical control structure used in diagnosis as proposed in this study is assumed to be of a moderate level of generality. It is general in the sense that it can be applied to more than one subdomain of paper making, as shown by the behavior of the transfer subjects. However, it is specifically tied to diagnosis, which may enable the recognition of this strategy as applicable to other diagnostic problems with similar problem features. Even if it may appear as a side-effect of having solved a large number of diagnostic problems, it may still turn out to be useful to try to teach this strategy, since it is assumed to greatly enhance problem solving performance, and would thus be a major contribution to the development of diagnostic skill.

References

Akin, O. (1980). *Models of architectural knowledge.* London: Pion.

Anderson, J.R. (1982). Acquisition of cognitive skill. *Psychological Review, 89,* 369-406.

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,* 192-210.

Anderson, J.R., Boyle, C.F., Farrell, R.G., & Reiser, B.J. (1984). Cognitive principles in the design of computer tutors. In *Proceedings of the Sixth Cognitive Science Society Conference.* Boulder, CO.

Anderson, J.R., Boyle, C.F., & Reiser, B.J. (1985). Intelligent tutoring systems. *Science, 228,* 159-175.

Anderson, J.R., Farrell, R.G., & Sauers, R. (1984). Learning to program in LISP. *Cognitive Science, 8,* 87-129.

Anderson, J.R., & Reiser, B.J. (1985). The LISP tutor. *Byte, 3,* 159-175.

Anzai, Y., & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review, 86,* 124-140.

Bainbridge, L. (1984). *Diagnostic skill in process operation.* Revised version of invited review paper presented at the international conference on occupational ergonomics.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error.* New York: John Wiley & Sons.

Bainbridge, L. (1988). Types of representation. In L.Goodstein, H. Andersen, & S. Olsen (Eds.), *Tasks, errors and mental models*. Basingstoke, Hampshire: Taylor & Francis Ltd.

Bakker, R.R., Hogenhuis, P.A., van Soest, D.C., & Mars, N.J.I. (1988). Diagnosis of technical systems based on structural models. In *Proceedings AI Toepassingen '88.*

Bakker, R.R., & Mars, N.J.I. (1989). *Diagnostische kennissystemen voor technische toepassingen.* Tutorial AI-Toepassingen 1989.

Bakker, R.R., Hogenhuis, P.A., Mars, N.J.I., & van Soest, D.C. (1989). Diagnosis of technical systems using design descriptions. In *Special Conference Second Generation Expert Systems and their Applications.* Avignon, France.

Barfield, W. (1986). Expert-novice differences for software: implications for problem-solving and knowledge acquisition. *Behaviour and Information Technology, 5*, 15-29.

Benjamins, V.R., & W.N.H. Jansweijer (1989). Modellen bij diagnostisch redeneren. In L. Siklossy & R.R. Bakker (Eds.), *Proceedings NAIC '89.* Schoonhoven, Academic Service, 1989.

Bennett, J. (1985). ROGET: A knowledge-based system for acquiring the conceptual structure of a diagnostic expert system. *Journal of Automated Reasoning, Vol. 1*, 49-74.

Bonar, J.G. (1985). *Mental models of programming loops.* Technical report, Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Boshuizen, H.P.A. (1989). *De ontwikkeling van medische expertise. Een cognitief-psychologische benadering.* PhD-thesis Rijksuniversiteit Limburg.

Breuker, J.A., & Wielinga, B.J. (1985). KADS: Structured knowledge acquisition for expert systems. In *Proceedings of Fifth International Workshop on Expert Systems and their Applications.* Avignon, France.

Breuker, J.A., & Wielinga, B,J. (1987). Use of models in the interpretation of verbal data. In A.L. Kidd (Ed.), *Knowledge acquisition for expert systems. A practical handbook.* New York: Plenum Press.

Breuker, J.A., Wielinga, B.J., v. Someren, M., de Hoog, R., Schreiber, G., de Greef, P., Bredeweg, B., Wielemaker, J., Billault, J., Davoodi, M., & Hayward, S. (1987). *Model-driven knowledge acquisition: Interpretation models.* Deliverable task A1, Esprit Project 1098.

Brigham, F., & Laios, L. (1975). Operator performance in the control of a laboratory process plant. *Ergonomics, 18,* 53-66.

Brown, A. (1976). *Qualitative knowledge, causal reasoning, and the localization of failures.* Technical report 362, MIT Artificial Intelligence Laboratory.

Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering, and understanding. In J.H. Flavell, & E.M. Markman (Eds.), *Handbook of child psychology.* Vol. 3: *Cognitive development.* New York: John Wiley & Sons.

Calkin, J.B., & Witham, G.S. (Eds.) (1957). *Modern pulp and paper making (third edition).* New York: Reinhold Publishing Corporation.

Card, S., Moran, T.P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Charness, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology, 33,* 1-16.

Chase, W.G., & Anders Ericsson, K. (1981). Skilled memory. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Chase, W.G., & Simon, H.A. (1973a). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing.* New York: Academic Press.

Chase, W.G., & Simon, H.A. (1973b). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chiesi, H., Spilich, G.J., & Voss, J.F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18,* 257-273.

Claessen, H.F.A., & Boshuizen, H.P.A. (1985). Recall of medical information by students and doctors. *Medical Education, 19,* 61-67.

Clancey, W.J. (1985). Heuristic classification. *Artificial Intelligence, 27,* 215-251.

Clancey, W.J. (1987). *Knowledge-based tutoring. The GUIDON program.* Cambridge, MA: MIT Press.

Crossman, E.R.F.W., & Cooke, J.E. (1974). Manual control of slow-response systems. In E. Edwards & F.P. Lees (Eds.), *The human operator in process control.* Londen: Taylor & Francis.

Cuney, X. (1979). Different levels of analysing process control tasks. *Ergonomics, 22,* 415-425.

Dague, P., Raiman, O., & Deves, P. (1987). Troubleshooting: When modeling is the difficulty. In *Proceedings of the 6th National Conference on Artificial Intelligence,* Seattle, WA.

Davis, R. (1984). *Diagnostic reasoning based on structure and behavior.* MIT Artificial Intelligence Laboratory, A.I. memo 739, 1984.

Davis, R., & Hamscher, W. (1988). Model-based reasoning: Troubleshooting. In H.E. Shrobe (Ed.), *Exploring artificial intelligence.* San Mateo: Morgan Kauffman Publishers Inc.

Edwards, E., & Lees, F.P. (1974). *The human operator in process control.* London: Taylor & Francis.

Egan, D.E., & Schwartz, B. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition, 7,* 149-158.

Elshout, J.J. (1983). Een beginner is méér dan iemand die het nog niet kan. In P. Drenth, W. Koops, J. Orlebeke, & R. Takens (Red.), *Psychologie in Nederland.* Amsterdam: Swets & Zeitlinger.

Engle, R.W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. *American Journal of Psychology, 91*, 673-689.

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

Ernst, G.W., & Newell, A. (1969). *GPS: A case study in generality and problem solving.* New York: Academic Press.

Feltovich, P.J (1981). *Knowledge based components of expertise in medical diagnosis.* Report No. PDS-2. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Feltovich, P.J., & Barrows, H.S. (1984). Issues of generality in medical problem solving. In H.G. Schmidt & M.L. de Volder (Eds.), *Tutorials in problem-based learning.* Assen: Van Gorcum.

Genesereth, M. (1984). The use of design descriptions in automated diagnosis. *Artificial Intelligence, 24*, 411-436.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 631-666.

Glaser, R., & Chi, M.T.H. (1988). Overview. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum.

Goldbeck, R.A., Bernstein, B.B., Hillix, W.A., & Marx, M.H. (1957). Application of the half-split technique to problem-solving tasks. *Journal of Experimental Psychology, 53*, 330-338.

Gott, S.P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E.Z. Rothkopf (ed.), *Review of Research in education, 15*. Washington, DC: AM. Educ. Res. Assoc.

Gott, S.P., & Pokorny, R. (1987). The training of experts for high-tech work environments. In *Proceedings of the Ninth Interservice/Industry Training Systems Conference.* Washington, DC.

Greeno, J.G. (1978). A study of problem solving. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 1.* Hillsdale, NJ: Lawrence Erlbaum.

Greeno, J.G., & Simon, H.A. (1988). Problem solving and reasoning. In R.C. Atkinson, R.J. Herrnstein, G. Lindzey & R.D. Luce (Eds.), *Stevens' handbook of experimental psychology. Volume 2: Learning and cognition (2nd edition).* New York: John Wiley & Sons.

Groot, A.D. de (1946). *Het denken van den schaker.* Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.

Groot, A.D. de (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving.* New York: John Wiley & Sons.

Hayes, J.R. (1985). Three problems in teaching general skills. In S.F. Chipman, J.W. Segal, & R. Glaser (Eds.), *Thinking and learning skills. Volume 2: Research and open questions.* Hillsdale, NJ: Lawrence Erlbaum.

Halasz, F.G., & Moran, T.P. (1983). Mental models and problem solving in using a calculator. In *Proceedings of CHI'83 Human Factors in Computing Systems.* New York, ACM.

Hamel, R. (1990). *Over het denken van de architect. Een cognitief psychologische beschrijving van het ontwerpproces bij architecten.* PhD-thesis, University of Amsterdam.

Hamscher, W.C. (1988). *Model-based troubleshooting of digital systems.* Technical report 1074, MIT Artificial Intelligence Lab.

Jansweijer, W.N.H., Benjamins, R., & Bredeweg, B. (1989). Diagnostic reasoning of the service engineer. Paper submitted to Conference on Second Generation Expert Systems, Avignon, 1989.

Jeffries, R., Turner, A.T., Polson, P.G., & Atwood, M.E. (1981). The processes involved in designing software. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

Kieras, D.E. (1982). *What people know about electronic devices: A descriptive study.* Technical report No. 12. UARZ/DP/TR-82/ONR-12. University of Arizona, Department of Psychology.

Kieras, D.E. (1987). *What mental model should be taught: Choosing instructional content for complex engineered systems.* University of Michigan, Technical Report No. 24.

Kieras, D.E., & Bovair, S. (1984). The role of a mental model in learning to operate a device. *Cognitive Science, 8,* 255-273.

Klahr, D., & Carver, S. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20,* 362-404.

de Kleer, J. (1986). *Local methods for localizing faults in electronic circuits.* Memo 394, MIT Artificial Intelligence Lab.

de Kleer, J., & Williams, B.C. (1986). Reasoning about multiple faults. In *Proceedings of the AAAI-86.*

de Kleer, J., & Williams, B.C. (1987). Diagnosing multiple faults. *Artificial Intelligence, 32,* 97-130.

Konst, L., Wielinga, B.J., Elshout, J.J., & Jansweijer, W.N.H. (1983). Semi-automated analysis of protocols from novices and experts solving physics problems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence.*

Kragt, H. (1983). *Operator tasks and annunciator systems. Studies in the process industry.* PhD-thesis Technical University Eindhoven.

Kragt, H., & Landeweerd, J.A. (1974). Mental skills in process control. In E. Edwards & F.P. Lees (Eds.), *The human operator in process control.* Londen: Taylor & Francis.

Kuipers, B.J., & Kassirer, J.P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8,* 363-385.

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.* Hillsdale, NJ: Lawrence Erlbaum.

Larkin, J.H., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.

Larkin, J.H., Reif, F., Carbonell, J., & Gugliotta, A. (1988). FERMI: A flexible expert reasoner with multi-domain inferencing. *Cognitive Science, 12,* 101-138.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing x-ray pictures. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum.

Long, W.J., Naimi, S., Criscitiello, M.G., & Jayes, R. (1986). Using a physiological model for prediction of therapy effects in heart disease. In *Computers in cardiology.* Cambridge, MA: MIT Press.

MacDonald, R.G., & Franklin, J.N. (1970). *Pulp and paper manufacture. Volume III: Papermaking and paperboard making.* New York: McGraw-Hill.

Mayer, R.E., & Greeno, J.G. (1972). Structural differences between learning outcomes produced by different instructional methods. *Journal of Educational Psychology, 63,* 165-173.

Mayer, R.E., Stiehl, C., & Greeno, J.G. (1975). Acquisition of understanding and skill in relation to subjects' preparation and meaningfulness of instruction. *Journal of Educational Psychology, 67,* 331-350.

McKeithen, K.B., Reitman, J.S., Rueter, H.H., & Hirtle, S.C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13,* 307-325.

Miyake, N., & Norman, D.A. (1979). To ask a question one must know enough to know what is known. *Journal of Verbal Learning and Verbal Behavior, 18*, 357-364.

Moraal, J. (1988). Validatie van simulatoren. In *Symposium boek: Simulatie: Een veelzijdig hulpmiddel bij training, nabootsing en onderzoek*. TNO.

Morris, P.E., Gruneberg, M.M., Sykes, R.N., & Merrick, A. (1981). Football knowledge and the acquisition of new results. *British Journal of Psychology, 72*, 479-483.

Morris, N.M., & Rouse, W.B. (1985). Review and evaluation of emprirical research in troubleshooting. *Human Factors, 27*, 503-530.

Murphy, M.D., & Puff, C.R. (1979). Free recall: Basic methodology and analyses. In C.R. Puff (Ed.), *Handbook of research methods in human memory and cognition*. New York: Academic Press.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Paige, J.M., & Simon, H.A. (1966). Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), *Problem solving*. New York: John Wiley & Sons.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: John Wiley & Sons.

Patel, V., Groen, G.J., & Arocha, J.F. (1990). Medical expertise as a function of task difficulty. *Memory and Cognition, 18*, 394-406.

Peterson, J.L. (1981). *Petri net theory and the modeling of systems*. Englewood Cliffs, NJ: Prentice-Hall.

Pirolli, P.L., & Anderson, J.R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology, 39*, 240-272.

Pople, H. (1982). Heuristic methods for imposing structure on ill-structured problems: The structure of medical diagnosis. In P. Szolovits (Ed.), *Artificial intelligence in medicine*, AAAS Selected Symposium 51, Boulder: Westview Press Inc.

Potter, N.R., & Thomas, D.L. (1976). *Evaluation of three types of technical data for troubleshooting*. Tech Rep. 76-74(3). Brooks, AFB, TX: Air Force Human Resources Laboratory.

Rasmussen, J. (1979). *On the structure of knowledge - a morphology of mental models in a man-machine system context.* Riso Report Riso-M-2192, Roskilde, Denmark: Riso National Laboratory.

Rasmussen, J. (1986). *Information processing and human-machine interaction. An approach to cognitive engineering*. Amsterdam: Elsevier Science Publishers.

Rasmussen, J., & Jensen, A. (1974). Mental procedures in real-life tasks: A case study of electronic troubleshooting. *Ergonomics*, *17*, 293-307.

Rasmussen, J., & Lind, M. (1981). Coping with complexity. In *European Annual Conference on Human Decision Making and Manual Control.* Delft: University of Technology.

Reitman, J.S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. *Cognitive Psychology*, *12*, 554-581.

Roenker, D.L., Thompson, C.P., & Brown, S.C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*, 45-48.

Rouse, W.B. (1982). A mixed fidelity approach to technical training. *Journal of Educational Technology Systems*, *11*, 103-115.

Rouse, W.B., & Hunt, R.M. (1984). Human problem solving in fault diagnosis tasks. In W.B. Rouse (Ed.), *Advances in man-machine systems research*. London: JAI Press.

Scarl, E., Jamieson, J.R., & Delaune, C.I. (1985). A fault detection and isolation method applied to liquid oxygen loading for the Space Shuttle. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, Los Angeles, CA.

Schaafstal, A.M. & Vos, A.M.A.M. (1988). De ontwikkeling van een kennissysteem voor operators in de papierindustrie. In *Proceedings AI-Toepassingen '88*.

Schaafstal, A.M. (1989). Kijk op procesinstallaties en kennis van storingen. In *Proceedings AI-Toepassingen '89*.

Schoenfeld, A.H. (1979). Explicit heuristic training as a variable in problem-solving performance. *Journal of Research in Mathematics Education*, *10*, 173-187.

Schraagen, J.M.C. (1989). *Verschillen in kennis tussen beginnende en gevorderde Damage Control Officieren*. IZF-rapport 1989-35.

Simon, H.A. (1980). Problem solving and education. In D.T. Tuma, & F. Reif (Eds.), *Problem solving and education: Issues in training and research*. Hillsdale, NJ: Erlbaum.

Simon, D.P., & Simon, H.A. (1978). Individual differences in solving physics problems. In R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum.

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

Sloboda, J.A. (1976). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology*, *28*, 1-16.

van Soest, D.C., Bakker, R.R., Hogenhuis, P.A., & Mars, N.J.I. (1988). Diagnosis of technical systems. In *Proceedings AI Toepassingen '88*.

Steels, L., & v.d. Velde, W. (1985). Learning in second generation expert systems. In J.S. Kowalik (Ed.), *Knowledge-based problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Stefik, M. (1981). Planning with constraints (MOLGEN: Part 1). *Artificial Intelligence, 16*, 111-140.

Szolovits, P. (Ed.) (1981). *Artificial intelligence in medicine.* Boulder: Westview Press, Inc.

Thorndyke, P.W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, *14*, 560-589.

VNP (1989). *Jaaroverzicht 1989.* Vereniging van nederlandse papier- en kartonfabrikanten.

Weghorst, P.M.G. (1989). *Een simulator voor de procesindustrie.* Master's thesis University of Twente (in dutch).

Weiser, M., & Shertz, J. (1983). Programming problem representation in novice and expert programmers. *International Journal of Man-Machine Studies*, *14*, 391-396.

Wielinga, B.J., & Breuker, J.A. (1984). Interpretation models for knowledge acquisition. In T. O'Shea (Ed.), *Advances in artificial intelligence.* Amsterdam: North-Holland.

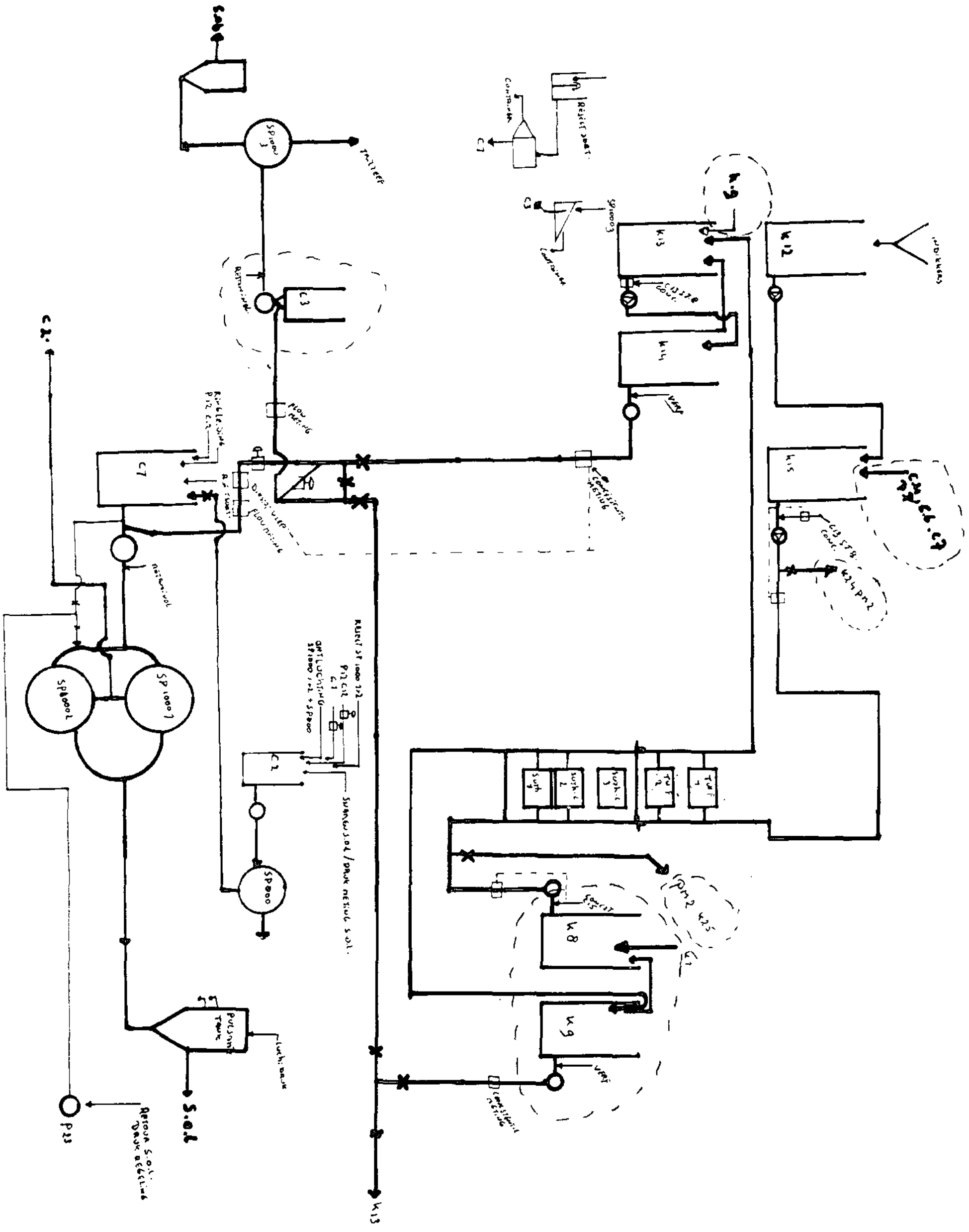
Wognum, P.M. (1990). *Explanation of automated reasoning: How and why?* PhD-thesis University of Twente.

Wognum, P.M., & Mars, N.J.I. (1989). Het belang van uitleg van redeneringen. In L. Siklossy & R.R. Bakker (Eds.), *Proceedings NAIC '89.* Schoonhoven: Academic Service.

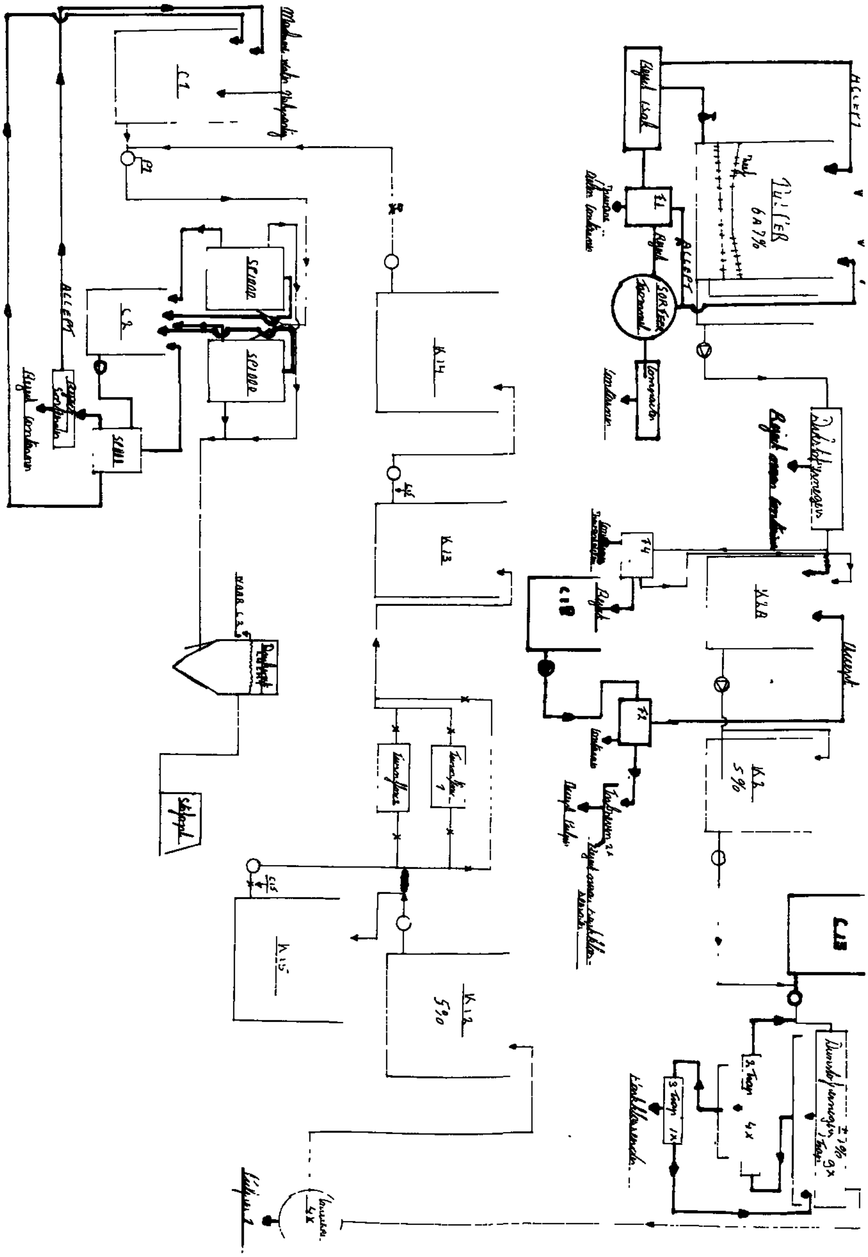
Appendix A

Examples: drawings of flowsheet

The following two pages contain, as representative examples, a drawing of the flowsheet by an expert and a novice operator.



Example drawing of the flowsheet by an expert operator. Bypasses have been circled.



Example drawing of the flowsheet by a novice operator.

Appendix B

Examples: Knowledge about process components

This appendix contains examples of answers given by subjects in experiment 4 reported in chapter Seven on various process components.

Fiberizer

Name:	Fiberizer
Purpose:	Cleaning, Separation
Controls:	Contains valves to control the amount of input and output
Input/Output:	Input: to be separated stock, input tangential Output: plastic goes to reject sorter good stock goes to buffer chest reject of reject sorter goes to container
Power source/ requirements:	Contains rotor to turn blade, sufficient rinsing water
External appearance:	Conical shape
Internal appearance:	Rotor, screen with cone-shaped holes, shaft, sluice for heavy pieces
External behavior:	Heavy pieces sluice opens once per time unit
Operation:	Inlet pressure should not not exceed 1.5 bar
How it works:	Stock is let in tangentially, the rotor causes it to turn. The light pieces stay inside. The heavy pieces are moved to the outside and go to the heavy pieces sluice and from there to container. Behind the blades of the rotor there is a screen with cone-shaped holes. The conic shape causes a pressure impulse: stock flows from high pressure to lower pressure.
Malfunctions:	Wire obstructed Inlet pressure too high

Reject sorter

Name:	Reject sorter
Purpose:	Cleaning, separation
Controls:	Amount of rinsing water can be regulated, flow of stock through different compartments can be regulated
Input/Output:	Input: from fiberizer Accept: to chest 22A Reject: to container
Power source/ requirements:	Contains rotor, needs rinsing water
External appearance:	Cylindrical
Internal appearance:	Shaft with blade system, wire at the bottom, several compartments, blades containing scrapers, sprinklers
External behavior:	Stock goes in at the left-hand-side, fibrous material leaves the reject sorter through the wire. Right-hand-side contains mainly plastic, which will be transported to container.
Operation:	The position of the guiding plates determines the throughput of stock. A more oblique position of the guiding plates forces the throughput to be faster.
How it works:	The to be cleaned stock comes in at the left-hand-side of the reject sorter and is gradually pushed through the different compartments to the right-hand-side. By the time it arrives there, all the fibrous material is separated from the plastic, and in the right-hand-side compartment almost only plastic is processed. By means of a fast spinning shaft with blades stock is pushed through the wire, which is positioned at the bottom of the reject sorter.
Malfunctions:	Clogging in a compartment Clogging of the wire Leakage Rotor disfunctioning

Reject drum

Name:	Reject drum
Purpose:	Sorting

Controls:	--
Input/Output:	Input: from Contaminex, reject of the pulper Reject: to container Accept: back to pulper
Power source/ requirements:	Sufficient rinsing water, speed not too high
External appearance:	Hollow cylinder
Internal appearance:	Reject drum is a wire, contains spiral and high pressure sprinklers
External behavior:	Reject of the pulper and fiberizers enters reject drum, plastic comes out, stock goes back to pulper
Operation:	--
How it works:	The reject drum spins. Stock falls to the bottom and is transported through the spiral. Fibrous material leaves the cylinder through the wire. Sprinklers are used to prevent the wire from clogging by means of cleaning and thinning fibrous material from the rejct. Reject stays inside and is transported through the spiral to the outside.
Malfunctions:	Inside clogging Clogging of inlet and outlet Sprinkler not working properly

Vertical screen

Name:	Vertical screen
Purpose:	Cleaning, separation
Controls:	Amount of reject, amount of input, differential pressure
Input/Output:	Input: -- Accept: to C1
Power source/ requirements:	Sufficient rinsing water, stable input pressure
External appearance:	Cylindrical
Internal appearance:	Wire with slots, impulse rotor or propellor, scraper
External behavior:	Final stock cleaning before the headbox. Input tangential
Operation:	Differential pressure should be kept within certain limits, by varying the amount of reject, and input
How it works:	Rotor moves in opposite direction compared to direction of stock. At the bottom there is a wire. Scraper turns

	around the wire to cut the fibers. The wire contains cone-shaped holes to establish a vacuum. Fibrous material is pushed through the wire. N.B. This actually turned out to be an improper description of how the working of the device.
Malfunctions:	Clogging Scraper to far apart from wire Insufficient amount of reject

Refiner

Name:	Refiner, Sutherland, Twinflow
Purpose:	Refining, in this paper mill cleaning by means of beating the dirt
Controls:	Oil pressure
Input/Output:	Input: long fibrous material from a buffer chest Output: more fibrillated material
Power source/ requirements:	Appropriate oil pressure, sufficient cooling water
External apperance:	Two beating blades
Internal appearance:	Compartment, outside are two beating blades.
External behavior:	The more oilpressure, the more refining. Lower degree of refining: beating of dirt. Higher degree of beating (more energy): beating of fibrous material.
Operation:	Lower degree of beating: beating of dirt. Higher degree of beating: beating of fibrous material. Beating is not done very often anymore.
How it works:	The Sutherland is a refiner which contains a compartment, in which the fibers are pushed against each other by means of beating blades. The closer the beating blades are to each other, the higher degree of beating.
Malfunctions:	Beating blades wrongly adjusted Amperage too high No or poor stock flow

Appendix C

Glossary of paper making terminology

Agitators

Type of auxiliary equipment, usually propellor like, in buffer chests or pulpers, used for blending and circulating stock or pulp.

Beating

See: refining

Buffer chest

Buffer chests, or stock chests, are large tanks, usually either cylindrical in shape and verical, or rectangular and vertical. Most chests are constructed of concrete.

Cleaning battery

A cleaning battery consists of several cleaners (for example, eight) and is a very important component of the process of thinstock cleaning. Each cleaner consists of a hollow cone section joined at the top by a cylindrical head or a cylindrical body with cone-shaped outlet connections. A stream of dilute suspension of the pulp is forced into the tangentially arranged inlet. The heavier particles are thrown to the outside, while the purified pulp is withdrawn from the center of the device which has a much lower pressure than that of the entering stream. Dirt and other foreign matter are separated from the paper stock by making use of differences in specific gravity of the fibers and dirt. Since some good fiber is always rejected with the dirt, several stages of fiber separation are necessary (for example, first, second, and third stage).

Consistency

The weight of fiber contained in a given weight of stock.

Contaminex

Device used in stock preparation for further cleaning of the reject of the pulper.

Conveyer belt

A conveyer belt is used to transport wood or waste paper into the pulper. If the conveyer belt is malfunctioning, the pulper is not feeded anymore, unless base material is fed in manually.

Cyclone

Element of centrifugal cleaning (thickstock and thinstock cleaning). Thickstock cleaning is one of the earlier cleaning processes in stock preparation, by means of which staples, iron elements, pieces of glass and so on are removed.

Dry broke

Paper that is torn of in the dry end of the paper machine. If it occurs, it should be removed from the cylinders to ensure a proper functioning of the dry end, and the prevention of more paper breaks. See also: wet broke.

Dryer section (dry end)

That part of the paper machine concerned with the removal of water from the sheet, mainly by evaporation. In drying, two basic physical processes are involved: heat transfer and mass transfer. Heat is transferred from some source, such as steam from cylinders, to the wet sheet in order to provide the energy required to drive the moisture from the sheet. The moisture evaporates and is then transferred from the sheet to the surrounding atmosphere by the mass transfer process.

Fiberizer

Cleaning device for light volume particles used in stock preparation.

Fourdrinier wire

A Fourdrinier is a very common type of paper machine with a Fourdrinier wire. A Fourdrinier wire is a long, endless wire, woven from synthetic fabrics, such as polyamide and polyesters.

Headbox

The headbox, by appearance, is similar to a large rectangular tank. The function of the headbox is to deliver the entire flow of water and pulp onto the wire, and thus, the headbox is the first part of the paper machine. There should be no undue pressures within the headbox which would create current and boiling action. The stock should approach the slice, and be delivered to the wire as a smooth sheet over the full width of the wire. The pulp fibers as they are laid on the wire should be drawn out and well separated and thoroughly mixed. If this delivery is not accomplished, the sheet will be uneven, will not dry properly and wil tend to break and tear during the manufacturing process. Thus, a proper functioning of the headbox is of the utmost importance in paper making.

Heavy pieces container

The heavy pieces container is part of the primary cleaning process of the pulper. Reject is moved to the container, accept is fed back into the pulper.

Machine chest

The last buffer chest for the headbox of the paper machine. Important chest, since level changes are directly reflected in the paper machine. If the stock flow to the machine chest is stopped for any reason, this will almost inevitably result in a halting of the paper machine.

Press fabric

Fabric used in the press section. Usually an endless fabric. The openness of the felt weave is very important for the ability to remove moisture from the paper.

Press section

Section of the paper machine. Sheet transfer from wire to press section is often combined with pressing, carried out by the pick-up press. The sheet entering the press from the wire can contain 80% water, depending on the machine speed and the grade of paper. Since it is extremely costly to remove water by steam drying, the efficiency of the press section is of critical importance to ensure economical operation. There are various kinds of press configurations for each type of paper and machine speed, such as suction presses, and extended nip presses.

Press pit

Part of the broke-handling system. Small chest located under the press section.

Pulp

The crude fiber material of one kind or another that is produced either mechanically or chemically from fibrous cellulose raw material or waste paper and from which, after suitable treatment, paper, paperboard and the like are made. Nowadays, an increasing use is made of waste paper as main ingredient for pulp.

Pulper

Pulpers are the first components in stock preparation, aimed at resuspending the fiber in water at the desired consistency and blending it with the other components, both fibrous and nonfibrous (types of paper and chemicals).

Reel

After the sheet leaves the papermachine, it must be put into a form that can be easily transported and handled, adapted to customer's needs. The first step in this process is to wind the paper on a reel.

Refining

Mechanical action performed on stock during stock preparation, in which the fibers are fibrillated. Refining is assumed to enhance fibrillation of the fibers, which is an important factor in the strength of the final sheet of paper. Refining is often used in the same manner as beating.

Reject sorter

Cleaning device for light small particles used in stock preparation.

Secondary pulper

Second stage of the pulping process. Consists of a second pulper which is usually smaller than the major one.

Sedifloat

A sedifloat system is used to separate water from fibrous material by means of flotation and sedimentation.

Selectifier

A selectifier is a vertical screen with a rotor. The stock flow may goe from outside to inside while the particles cannot pass through are removed from the outside of the screen and are pushed through a reject outlet.

Sheet formation

Sheet formation is an important factor in obtaining a good sheet of paper. It consists of a homogeneous distribution of fibers over the sheet. Sheet formation is accomplished early on in the paper making process (in the wet end).

Sorting drum

Part of the cleaning process in stock preparation.

Stock preparation

Those operations and mechanial treatments necessary to make fibers suitable for forming into a sheet of paper. Examples of operations in stock preparation are repulping and blending of pulps of different types, and the addition of various chemicals and fillers, such as clay. Examples of mechanical treatment are beating and refining.

Thickener

Type of separator used to enable reuse of fibrous material by means of separation of fibrous material from water. Principal aim of thickeners is to obtain stock with a higher consistency. Thickeners are part of stock preparation.

Thickstock cleaning

See: cyclone.

Thinstock cleaning

See: cleaning battery.

Vertical screen

See: selectifier

Wet broke

The term broke refers to partly or completely manufactured paper that does not leave the machine room as paper or board that may be sold. Wet broke is broke made before the paper reaches the dryer section. On large high-speed machines, broke can accumulate at a high rate. It is necessary, therefore, to have a broke-handling system that will dispose of the broke as fast as it is produced. Such a broke-handling system includes several pits, such as a hog pit, a press pit, and a couch pit.

Wet section (wet end)

The wet end is considered a combination of headbox, wire section and press section. The wet section is considered a combination of headbox and wire section.

White water

White water is water carrying fibrous material, chemicals, and heat. It is reused as optimal as possible to keep the fibrous material in the process, to avoid pollution, and to reduce the costs of clarification of the plant effluent. The opposite of white water is fresh water, water taken in from the outside (for example a river).

Winder

Winders are used to slit and trim the mother roll taken from the paper machine into the size required by the converting department or the costumer. In the mill studied the winder is the completion of the paper making process.

Winder snap-offs

Winder snap-offs are paper breaks occurring at the winder, usually caused by either a problem in the winder itself, or an (un)noticed problem at the paper machine.

Wire screen

See: Fourdrinier wire.