ERGONOMICS, 1996, VOL. 39, NO. 7, 966–979

Individual differences and the effects of an information aid in performance of a fault diagnosis task

JEROEN G. W. RAAIJMAKERS* and WOUTER W. VERDUYN†

*TNO Human Factors Research Institute, Soesterberg, The Netherlands
†Royal Netherlands Navy, The Hague, The Netherlands

Keywords: Decision aiding; Diagnosis; Problem solving; System knowledge.

Two experiments are reported that investigated the performance of operators in a fault diagnosis task. Naval engineers working in the Ship's Control Centre (SCC) on board of frigates of the Royal Netherlands Navy were asked to solve a number of unfamiliar fault problems. In the first experiment, it was shown that there was a high correlation (0·85) between the number of problems correctly solved and the availability of system knowledge. In addition there was an effect of the level of formal education. In a second experiment it was shown that the provision of a simple help facility that compensates for a lack of system knowledge leads to a substantial increase in the number of problems that were correctly solved. The percentage of correct solutions increased from 62 to 89%. This increase was the same for operators with either an electrical engineering or a mechanical engineering background although the former group performed slightly better than the latter. These results show that the conclusion of Morris and Rouse (1985a) that instruction in the theoretical principles on which a system is based is less effective than the training of diagnostic procedures, underestimates the importance of system knowledge for the solution of fault diagnosis problems.

1. Introduction

Owing to the increasing complexity of modern technical systems, the task demands of the human operator have gradually changed from physical activities to activities of a more cognitive nature. One of the most important tasks for operators in a technical installation is that of fault management or troubleshooting: given that there is a system failure the damage to the installation and the environment has to be controlled, the cause for the malfunction must be quickly determined and appropriate action must be taken to correct the malfunction.

In this article two experiments will be described that investigate the diagnostic skill of operators that is involved in the second mentioned aspect of fault management, i.e. the determination of the cause for the system failure based on an analysis of the observed symptoms. Much research has shown that such diagnostic tasks are particularly difficult for human operators and quite susceptible to error (Rasmussen 1981, 1986, Morris and Rouse 1985a), especially when the problem is a novel, non-routine one. In such cases, the operator cannot rely on rules based on past experience but has to use knowledge about the system in an effort to solve the problem by reasoning. Given that such tasks are quite difficult and error-prone, a natural question is to what extent operators can be supported and whether such aids are indeed effective.

The experiments to be reported concern diagnostic performance in a real-life technical environment. Naval engineers working in the Ship's Control Centre (SCC) on board of frigates of the Royal Netherlands Navy were asked to solve a number of unfamiliar fault diagnosis problems. Unfamiliar problems were presented in order to avoid the more or less automatic application of well-practiced standard solutions. Hence, the task required (in Rasmussen's terminology, see Rasmussen 1983) *knowledge-based* behaviour. One practical concern that motivated the present study was the introduction of a new type of frigate (the so-called Multi-purpose or M-frigate). This frigate has a completely redesigned SCC in which all technical information is presented on a set of graphical and alphanumeric displays. The question then is whether users will in fact be able to solve diagnostic problems using the information available.

Pilot research (Boer 1992) suggested that the answer to that question was negative: subjects could not provide a correct solution in approximately 60% of the presented problems, all involving novel faults. A re-analysis of these data, in which a more lenient criterion was used (a solution was considered to be correct if it differed only in specific details from the normative solution but not in the determination of the general area where the problem was located), showed that 40% was still judged to be incorrect.

The first experiment that was presented aimed to determine which factors were responsible for the poor diagnostic performance. As in Boer (1992) subjects were presented with a number of scenarios or problems, each scenario including one or more alarm messages produced by a specific system failure. The task of the subjects was the localization of the defective part of the system. In addition, the subjects were given a test in which a number of questions had to be answered concerning the functional aspects of the system. This test was given in order to be able to correlate diagnostic performance with a measure for system knowledge.

2. Experiment 1

2.1. *Method*

2.1.1. *Subjects*: Eight subjects participated in this experiment. The subjects were selected on the basis of their experience with the M-frigate. No subject had extensive experience with the specific aspects of the system because of the recent introduction of these frigates. However, the problems given to the subjects did not require knowledge specific to the M-frigate, but could be solved based on (a) general knowledge concerning similar ships with which they did have extensive experience, and (b) the skills necessary to interpret the system layout that was available to them.

The subjects differed in terms of number of years of experience in their present or similar functions, as well as in the nature of their function (i.e. their rank or function level). In order to get some information concerning the relation between diagnostic behaviour and function level, the different functions were categorized into three levels. The first level (three subjects) involved lower operator functions (rank of seaman). The second level (three subjects) involved functions requiring the rank of petty officer. The remaining two subjects held officer positions.

2.1.2. *Materials*: Eight scenarios were presented to the subjects. The scenarios all dealt with problems involving the ship's propulsion system. In the M-frigate the propulsion system consists of two nearly identical sets of engines, each driving one propeller (one port and one starboard). Each of the two installations consists of a number of components, i.e. the engines themselves plus a number of auxiliary systems that control fuel, oil and cooling. One of the tasks of an operator in a SCC is to monitor and control correct functioning of the propulsion system. To this end the operator has available an interface that displays a large number of system parameters (e.g. temperature and pressure information provided by sensors at particular locations) as well as a message display that automatically generates alarm messages whenever one of the system parameters falls below or above a criterion value.

All of the eight problems could be solved using general technical principles (e.g. pressure decreases when there is a leakage, when there is a stoppage before the pressure sensor, or when the pumps do not have sufficient capacity), and did not require knowledge specific to the particular type of ship. They all involved 'normal' faults such as a leakage, operator errors (incorrect settings), a cooler that is malfunctioning, or a dirty filter. Prior to the presentation of each problem, subjects were given information about the general operating conditions of the ship (number and type of engines turned on, and speed).

These scenarios varied in complexity and difficulty but could all be solved within a short period of time using the available information. This information was presented on paper in a form that resembled the actual screens in the SCC. The sensor values were not printed but replaced by common symbols such as T (temperature), P (pressure) and C (volume). Whenever subjects wanted the actual value, they had to ask the experimenter, giving also the reason why this value would be useful. The subjects were allowed to write the information down. This procedure was used in order to obtain additional data on sequential aspects of the problem-solving process. These data will not be analysed in this paper.

In order to get a measure of the amount of system knowledge of the subjects, a short test involving 20 questions was given. These questions involved various aspects of the propulsion system and did not require knowledge of aspects that are specific to the M-frigate. Each question was printed on a separate page. The questions were based on single steps in a normative solution developed by experts for the various scenarios used in the experiment. For example, how correct functioning of a part of the system should be checked, what symptoms can be seen when a certain part of the system is defective, or what are the possible causes for a given symptom. Some of the questions were split into a few sub-questions that were scored separately. The maximum score that could be obtained in this test was 39.

2.1.3. *Procedure*: Prior to the experiment, the subjects were asked to fill out a form giving information concerning age, operational experience and present job. Next, they were briefed concerning the basic purpose of the experiment and the task that they were required to perform. An example was given to illustrate the way in which information would be presented. For each scenario, information describing the general situation and the alarms was given on a separate sheet. Any other information that was deemed necessary had to be requested from the experimenter. Subjects were asked to indicate why they needed the information. This was written down by the experimenter and also recorded on tape.

Following the instruction, subjects were given the eight fault scenarios. In order to control for possible order effects, the order of the scenarios was balanced over subjects using a latin square design in which each scenario was followed and preceded by each other scenario exactly once.

On average, the subjects needed about 1 h to solve all eight scenarios. After they had finished, they were given a booklet with the knowledge test questions. On the first page of that booklet, a short instruction was given. The subjects were asked to give short answers to each question, in the order presented and without turning back to previous pages. It was stressed that there were no trick questions and that they were not required to give explanations.

This test took about 30 min. After they had finished, the subjects were given an explanation of the correct solution to each of the previously presented fault scenarios.

2.2. *Results and discussion*

The subjects' solutions for the eight scenarios were rated by two independent judges using the known normative solutions. For each scenario the quality of the diagnosis was rated on a 5-point scale:

1 = no diagnosis or diagnosis incorrect;
2 = diagnosis incorrect with respect to the main problem; sub-problem correctly diagnosed but relation with main problem not recognized;
3 = problem correctly localized but wrong component or explanation given;
4 = correct diagnosis but not localized in sufficient detail; and
5 = fully correct diagnosis.

Using a strict criterion for correctness (i.e. a score of 5), 31% of the problems (20 out of 64) were classified as being correct. This result agrees well with results from the pilot experiment in which the percentage classified as being correct was 34%. In that experiment, the subjects did not have any specific experience with this type of frigate, while in this experiment the subjects were well acquainted with the new frigate. Thus, the problems were indeed such that they did not require specific knowledge of this ship but only general technical knowledge.

It might be the case that the large percentage of errors was due to the strict criterion used. Perhaps most of the solutions were such that they were only incorrect in a minor detail. For this reason, the solutions were scored using a lenient criterion, i.e. a solution was considered correct if it only differed in specific details from the normative solution but not in the determination of the general area where the problem was located (score 3 or higher). Using this lenient criterion it was found that 62% were correct (40 out of 64). Thus, even though performance was indeed better using this criterion, still 38% of the problems were not solved correctly.

The most important question in this experiment concerns the relation between an operator's system knowledge and the quality of the diagnostic performance. The amount of system knowledge was determined using the total summed score on the knowledge test.

Table 1 gives for each subject the knowledge test score as well as the diagnostic performance score (number of problems solved correctly using the lenient criterion), the number of years experience in this or similar positions, and the level of their present functions (using the three-category classification scheme).

The mean test score was 26·1 (= 67% of the maximum score). As is evident in table 1, there is a strong relationship between the level of system knowledge and the quality of the diagnostic performance. The Pearson product-moment correlation between these two variables is ·847 ($F(1,6) = 15{\cdot}3$, $p < {\cdot}01$). Thus, a large part of the differences between subjects in their performance on this task can be explained by differences in system knowledge (this explains 72% of the observed variance).

A possible artefact of the present procedure that might explain this relation is that the test questions were always given after the subjects had solved the fault problems. Certain information that was required for the solution of the problems was also

Table 1. Experience, function level, diagnostic performance, and score on test for system knowledge.

Subject	Experience (years)	Function level	Diagnostic performance	Test score
1	7·5	2	4	21
2	3	1	4	22
5	11	1	4	25
3	7	2	5	28
8	1·5	3	5	25
4	7	1	6	32
6	11·5	2	6	29
7	6	3	6	27

required in the knowledge test. If subjects learn during problem solving, this might have a confounding effect on the observed correlation. However, a further analysis in which for each problem a separate score was computed using only those questions that were relevant for that problem, showed only a minor correlation (·316) between these scores and performance on that problem. Thus, it may be concluded that the observed correlation between the overall test score and diagnostic performance is not due to learning during diagnosis.

The remainder of the variance can be explained for the most part by differences in function level. The multiple correlation between diagnostic performance on the one hand and system knowledge and function level on the other, is ·94 ($F(2,5) = 18{\cdot}65$, $p < {\cdot}01$). This means that 88% of the variance (84% after adjustment for shrinkage, i.e. the inflation of the multiple correlation due to the small number of observations) may be attributed to differences in system knowledge and function level. By itself, there is not a strong relation between function level and performance: the correlation is only ·37. The relationship only shows up after differences in system knowledge have been partialled out: the partial correlation is ·76.

What factor is responsible for this correlation? One explanation is that function level correlated with general intelligence. That is, function level is correlated with the level of previous education and this is presumably related to intelligence. Quite a few studies in a more educational context have shown that more intelligent subjects make use of better 'meta-cognitive strategies' (Elshout and Veenman 1992, Veenman and Elshout 1991, Veenman 1993), i.e. general strategies that are used in problem solving (such as a more hypothesis-oriented approach, better 'bookkeeping', etc.) and that have been shown to have a positive effect on performance. Thus, the positive correlation between function level and performance might be interpreted as reflecting the fact that higher level subjects are better at employing the meta-cognitive strategies that are required for the solution of non-routine-like problems that draw heavily on general system knowledge.

Another possible explanation is that the correlation is due to the fact that officers, being primarily responsible for fault management on ships, have had more opportunities and more training to use meta-cognitive strategies than personnel from lower ranks. Hence, their performance might be better owing to more experience in this type of knowledge-based problem.

Whatever the reason for the correlation, it should be kept in mind that the correlation by itself is not very strong and only becomes evident after differences in system knowledge have been partialled out. However, these results do conform to recent views on diagnostic skills that assume that two types of knowledge are relevant for solving diagnostic problems in technical systems: knowledge of the system and knowledge concerning effective problem-solving strategies (Schauble *et al.* 1992).

There was no relation between diagnostic performance and the number of years of experience ($r = \cdot 13$, $p > \cdot 5$). Even when differences in function level are taken into account (in this group of subjects there is a negative relation between function level and number of years of experience), there is no relation between experience and the quality of problem solving. Apparently, the knowledge necessary for solving these problems is not simply acquired from experience. This lack of an effect of number of years of experience is probably related to the nature of the problems presented, which could not be solved with acquired skills but had to be dealt with on a knowledge-based level.

The present results clearly show that performance in this diagnostic task is highly correlated with the system knowledge that the operator possesses, the knowledge concerning the functional relationships between system parts. Based on these results one might conclude that this aspect should be given more emphasis in the training of operators.

Superficially, this conclusion seems to be in conflict with the much quoted assertion of Morris and Rouse (1985a,b) that instruction in the theoretical principles on which a system is based is less effective than the training of diagnostic procedures. Morris and Rouse (1985b) compared four training conditions that differed in the amount and type of training. The results showed that although the groups did not differ in their ability to detect and repair faults (including novel faults), those subjects that were given training in procedures controlled the experimental system in a more stable manner. Other research also showed that instruction in the abstract principles of a system is not an effective way to produce good troubleshooters.

However, this conclusion should not be overgeneralized. Morris and Rouse (1985a) also describe other research in which training in the use of context-specific knowledge did lead to better performance. Recent research also shows that training in system knowledge is not always ineffective. For example, Patrick and Haines (1988) compared the effect of two types of training materials, one involving a technical story and the other a set of diagnostic heuristics. They observed a positive effect of instruction in system knowledge (the technical story condition), even in situations where the layout of the test system differed from the layout of the training system.

As argued by Patrick and Haines (1988), it is important to use a more exact definition of 'theoretical' training or 'system knowledge'. Some types of training, those that put an emphasis on abstract principles without training in how to apply this to solve problems, may be less effective than a type of instruction that is directed at insight in the relevant variables and components of the system. What seems most important is instruction in the functional dependencies of the system. For example, how is a drop of pressure in one part of the system related to symptoms elsewhere in the system, or what are the consequences of a fault in a particular sub-system? This is the type of knowledge that we have called 'system knowledge' and that was shown to be highly correlated with the quality of the diagnostic performance.

A second problem with a simple application of the hypothesis of Morris and Rouse (1985a) is that there are not standard rules or procedures available for non-routine faults (as were used in the present study). In such cases, the operator can only make use of general knowledge of the system and general diagnostic strategies.

An improved availability of system knowledge can be obtained not only through training but also be a design of the support system such that the necessary information can be easily retrieved when it is needed. This might be done by giving the operator information concerning the relevant causal relations that might lead to a specific alarm. In the next experiment a type of diagnostic support that is based on this idea will be examined.

In conclusion, the results of the first experiment provided evidence (albeit of a correlational nature) for the hypothesis that lack of system knowledge was a main cause for poor diagnostic performance. On the basis of these results, a particular type of information aid was designed and evaluated in a second experiment. This diagnostic aid provided the subjects with information concerning the possible causes for an alarm as well as advice about which information might be consulted in order to test the various hypotheses concerning the cause of system failure. The aid therefore could compensate for the lack of relevant system knowledge that was hypothesized to be a major determinant of diagnostic performance. As discussed by Yoon and Hammer (1988), such an information aid has a number of advantages over other aiding approaches such as completely automated systems since it keeps the operator in control.

3. Experiment 2

The results of the previous experiment clearly showed that performance in this diagnostic task was far from adequate and that a lack of system knowledge was a major contributing factor. We therefore tried to design some sort of on-line help system that might improve performance.

There are several ways in which help might be provided (see Yoon and Hammer 1988, for a review). The major types are: *automatic diagnostic systems* in which the aiding system essentially takes over most of the work; *expert systems* that give advice based on rules and procedures derived from experts; *model-based systems* in which advice is given based on a dynamic model of the system; and, *information aids* that do not solve the diagnostic problem but provide the operator with relevant information. As argued by Yoon and Hammer (1988) there are a number of advantages to information aiding, the most important one being that it keeps the operator in the diagnosis and control task loop. In addition, information aiding is often easier to implement and much cheaper than more technologically advanced systems. For these reasons we decided in favour of a help system based on information aiding.

The system that was designed provided assistance in the interpretation of the alarm messages that were presented by the system whenever a deviation was detected in one of the parameters that were constantly monitored by the system (e.g. pressure or temperature in various parts of the system). The help system was such that when the operator moved the cursor to a particular alarm message and pressed the space-bar, help information would be presented in the bottom half of the screen. The help information consisted of a listing of possible causes for that alarm (hypotheses that should be tested during diagnosis) as well as the type of information that should be checked in order to test that hypothesis. See Verduyn (1993) for a complete description of all help screens for each of the scenarios that was used in this experiment.

Leakage hydraulic oil SB	53 level main compartment SB
	53 level auxiliary compartment SB
Insufficient capacity hydraulic pump SB	53 electric hydraulic pump SB
Blockage before hydraulic pressure sensor SB	53 pressure difference filters SB

Figure 1. Example of help screen. SB: starboard. Note that in the actual help screen, various terms have been abbreviated (e.g. hydr for hydraulic, aux for auxiliary).

In figure 1 an example is given of the type of help that was provided (in this case for the alarm: low pressure hydraulic system starboard). The first column lists one or more hypotheses. These could be refined in a second column (not present in this example). The last column indicates which sensor information on which mimic might be used to test the hypotheses. Thus, in this example, to test the hypothesis of oil leakage, the operator should check the levels of the main and auxiliary compartments on the mimic screen 53. It should perhaps be noted that the help system did not eliminate the need for diagnostic reasoning, i.e. subjects still had to interpret and combine the results from various observations in order to arrive at the correct diagnosis.

The help system did not include compensatory actions since this experiment focused on the diagnostic aspect of fault management and the subjects did not have to carry out emergency measures (e.g. reducing speed to avoid overheating the engine). In an actual implementation of the help system, this feature would have to be included.

The prototype design for the diagnostic aid was presented for evaluation to eight experienced naval officers (officers and chief petty officers). Their comments were incorporated in the final design.

3.1. *Method*

3.1.1. *Subjects*: For practical reasons it was not possible to use only subjects that had prior experience in the Ship's Control Centre of the new M-frigate. It was therefore decided to use a homogeneous group of subjects that all had similar levels of experience on board a similar type of frigate. All subjects had experience in the type of tasks that were used in the experiment.

Sixteen subjects participated in the experiment. They all were employed at the training centre of the Royal Netherlands Navy in Amsterdam. Eight subjects had an electrical engineering and eight had a mechanical engineering background. This does not affect the amount of experience in the type of task considered here but it does affect the other duties that have to be performed.

None of the subjects had any detailed knowledge of the new M-frigate prior to the experiment. All however had extensive experience with this type of propulsion system in general and had ample experience in solving diagnostic problems in technical systems.

3.1.2. *Materials*: A simulation system at the TNO Human Factors Research Institute was used. This set-up consists of three computer displays (two 19" graphic displays and one text terminal), a tablet and a keyboard as well as the main simulation computer (a SUN SparcStation 1).

The text terminal was positioned at the right-hand side and served as the display for alarms. The two other displays were used to present the mimics (14 in total) corresponding to the propulsion system. The tablet could be used in combination with a pen to request the display of a particular mimic. In addition, the simulation system allows for changes in speed and start/stop commands but this was not used in the present experiment.

The keyboard was used to interact with the help system. Using the cursor controls, the cursor could be placed on one of the alarms. Upon pressing the space-bar help information would be presented on the bottom half of the alarm display.

The same eight scenarios that were used in the first experiment were used. In addition, subjects were given three similar scenarios during the training phase prior to the experiment. Generally speaking, the interpretation of system alarms depends on the current context of system operation, i.e. how fast the ship is going and the types of engines that are operating. For simplicity, only two types of general operating conditions were used. In this way, subjects do not have to spend a large amount of time building a mental representation of the ship's condition (including the normal values for system parameters that depend on the current operating condition).

3.1.3. *Procedure*: Several days before the experiment the subjects were given a 90-min introduction on the purpose and the set-up of the experiment. At that time, they also visited a demonstration version of (part of) the SCC of the M-frigate where special attention was paid to the mimics for the propulsion system. Each subject was given a booklet in which background information about the experiment and a description of the propulsion system of the M-frigate was given using the mimics. Subjects were asked to reread the information prior to the experiment.

Participation in the experiment took 1 day. In the morning session of 2·5 h they were acquainted with the system. The experiment itself took place in the afternoon and also lasted about 2·5 h. In the morning session, the subjects were accustomed to the interface of the experimental system, and were given information about the propulsion system and the way in which the fault problems had to be dealt with.

First all the mimics were discussed. Then for each of the two operating conditions that were used in the experiment, the values of the most important system parameters were recorded on paper. These could be used in the experiment since the subjects were not expected to memorize all the values in the short time available.

Next the subjects were given a written instruction to prepare them for the three training problems. The subjects were informed that they were on their own and had to solve all problems using the information that was available. During the second training problem the use of the help system was explained and in the third training problem the subjects were allowed to use the help.

In the afternoon session they were presented with the eight experimental problems, four in which the help was not available and four in which it was. Before each problem scenario a brief written description was given of the ship's situation. The subjects were allowed to localize the cause for the fault using any procedure that they would use in real situations, except that they were required to work as fast and as accurately as possible (i.e. a typical speed instruction). When the subject indicated that he knew the cause of the fault or made clear that he could not continue any further, the simulation was stopped by the experimenter. The experimenter then noted the diagnosis that was given (asking for explanation if necessary) and the time that was used (measured from the time of the first alarm). The subjects were also asked to indicate their confidence in their solution to the problem.

For half of the subjects the diagnostic help was available for the first four problems, for the other half it was available for the last four problems. The order of the problems was randomized for each subject using a latin square design. Thus each problem was presented twice in each position, once with and once without the help available.

3.2. *Results and discussion*

3.2.1. *Quality of diagnostic performance*: The most important aim of this experiment was to see whether the addition of the help system would increase the percentage of correct diagnoses above the 60% level that was obtained in previous experiments. For each scenario the quality of the diagnosis was rated on the same 5-point scale as used in the previous experiment. As before, a strictly correct solution was defined as a solution with a score of 5. Using lenient scoring a solution with a score of 3 or higher was considered to be correct.

The results have been analysed using a repeated-measures analysis of variance with background (electrical versus mechanical engineering) as between-subjects factor and scenario and availability of help as within-subjects factors.

Table 2 gives the frequency distribution of the quality scores divided according to whether help was or was not available. The availability of the help system led to a substantial increase in the quality of the diagnostic performance. This was confirmed by the statistical analysis ($F(1,14) = 9{\cdot}49$, $p < {\cdot}01$).

Using a strict criterion (a score of 5), 16% of the diagnoses without help and 31% with help were correct. Using a lenient scoring criterion as discussed above (scores 3, 4, or 5), 62% of diagnoses were correct without help and 89% with help. (Subsequently, the lenient scoring system will be reported unless otherwise specified.) The percentages correct without the availability of the help system are close to the results obtained in experiment 1 in which the same problem scenarios were used.

The help system was used in 51 of the 64 problems in which it was available. Of the 13 cases in which it was not used, only one was incorrect (a score of 2). Of the

Table 2. Quality of diagnostic performance with and without help available.

	Help not available	Help available
Quality score	Frequency (%)	Frequency (%)
1	9 (14)	1 (2)
2	15 (24)	6 (9)
3	15 (23)	19 (30)
4	15 (23)	18 (28)
5	10 (16)	20 (31)

Table 3. Percentage of problems correctly solved (lenient criterion) with and without help available for subjects with an electrical or mechanical engineering background.

Background	Help not available	Help available
Electrical	66	94
Mechanical	59	84

remaining 12, two were given a score of 3, five a score of 4 and five a score of 5. Thus, it may be concluded that in general the decision not to use the help was justified.

Table 3 gives the percentage of problems correctly solved, expressed according to the background of the subjects (electrical or mechanical engineering). Both groups show a similar increase in performance. There was a small but significant main effect of the background of the subjects ($F(1,14) = 5{\cdot}05$, $p < {\cdot}05$): both with and without help, the subjects with an electrical engineering background score slightly better than those with a mechanical engineering background.

This difference may be due to the fact that mechanical engineers are more accustomed to use their perceptual senses in searching for causes of failures. Electrical engineers on the other hand more usually have to rely on a more abstract mode of problem solving (as is required for the present problems). It was originally thought that mechanical engineers would profit more from the help system, since that would force a more deductive, hypothesis-oriented approach. However, the increase due to the help system is about equal in both groups: there was no interaction between background and availability of help ($F(1,14) = 0{\cdot}19, p > {\cdot}05$). Apparently then, either the difference between the subjects is not so much due to a problem solving approach or the help system does not eliminate such differences. The first possibility is supported by the results of experiment 1 in which it was shown that the individual differences were largely due to differences in relevant system knowledge. Although the help system provides information that may serve to retrieve such information from memory, it does not fully compensate for any lack thereof (as is clear from the fact that performance is not perfect with the help available).

A separate analysis was performed to determine whether the order in which help was given (first four problems or last four problems) had any effect. There was no main effect of order ($F(1,14) = 0{\cdot}26$, $p > {\cdot}05$) nor an interaction of order and availability of help ($F(1,14) = 0{\cdot}39$, $p > {\cdot}05$).

3.2.2. *Response time*: In addition to its effect on accuracy, it was also determined whether the use of the help system affects the mean time that is needed to solve the problems (of course there are large differences between the various scenarios but these are to be expected). Without the help system, the mean time for solution was 228 s. With the availability of the help system there is a small but non-significant increase of 9% to 248 s. Neither the background of the subjects nor the availability of help or their interaction had a statistically significant effect on the solution time (in all cases, $p > {\cdot}25$).

The moment in time when the help system was first consulted was looked at also. There was no correlation between this variable and the quality of performance. There was a simple relation between the time at which help was first requested and the total solution time: the total solution time (T_t) was approximately equal to the time of the first request (T_1) plus 3 min: $T_t = 1{\cdot}01 T_1 + 182$. The linear correlation coefficient was ·58.

Prior to the experiment it was hypothesized that a relatively late request for help would lead to a reduction in the remaining solution time. This might be due to the fact that the subject would already have collected a substantial amount of information before help was requested. This expectation was not confirmed. It turns out that the subjects do not restrict themselves to those hypotheses (mentioned in the help) that they have not yet considered. Apparently, they have forgotten that they have already checked them or they do not recall why they had been rejected. Another possibility is that those subjects that ask for help relatively late are confused by the symptoms and their initial search has not led to any substantial progress.

3.2.3. *Confidence*: After they had provided a solution, the subjects were also asked to indicate their confidence in its correctness. This was scored on a 5-point scale. For both groups of subjects there was a mean increase in confidence with the availability of help of 0·15. Analysis of variance of the confidence scores showed however that this difference was not significant ($F(1,14) = 0{\cdot}52$, $p > {\cdot}25$).

Subjects with an electrical engineering background were slightly more confident than subjects with a mechanical engineering background ($F(1,14) = 4{\cdot}78$, $p < {\cdot}05$). This agrees with the finding that their solutions are also slightly better on average. There was no interaction effect of availability of help with background ($F(1,14) = 0{\cdot}02$, $p > {\cdot}25$).

4. General discussion

The results of the present experiments show that performance in a task in which unfamiliar system faults have to be diagnosed is mainly a function of relevant system knowledge. There may also be an effect of working method (more or less systematic) but this factor does not nearly explain as much of the variance as the availability of system knowledge.

Providing the subjects with access to a simple help system that gives a brief summary of the possible causes for a system failure and how that might be checked, leads to a substantial increase in the number of problems that are correctly solved, without leading to a concomitant increase in solution time.

From a purely applied point of view, such an increase is of course very important, since it shows that operator performance may be improved quite dramatically in a simple way that does not involve a large investment. The present system led to a decrease in the number of incorrect diagnoses from 38 to 11%. Compared to automatic diagnostic systems the present help system is not only much cheaper but also does not have the disadvantage of a decreasing involvement of the operator in the diagnostic process.

The present help system gives the operator the opportunity to ask for additional information concerning specific alarms. The help system lists a number of probable causes for the selected alarm and also indicates which other information should be checked in order to test that hypothesis. Although much of the presented information will be known, its advantage is that it is immediately available in stressful, high pressure situations. Moreover, access to the help system is easy and does not involve any kind of searching.

A more general result of the present research is that there seem to be limitations to the conclusion of Morris and Rouse (1985a) that instruction in the theoretical principles on which a system is based is less effective than the training of diagnostic procedures. The present results show that lack of relevant system knowledge is a main determinant of performance in non-routine situations and that providing subjects with such information leads to a substantial increase in performance. This conclusion agrees with the results of Patrick and Haines (1988) who also showed a positive effect of instruction in system knowledge. It is concluded that a distinction should be made between theoretical knowledge that involves only the general theoretical principles that underlie system performance and system knowledge that involves knowledge of functional relations within the system, the ability to reason about the system, to draw inferences, and to list possible causes for system failures.

In this experiment there was a difference in diagnostic accuracy between operators with an electrical and a mechanical engineering background, even though the problems were mainly of a mechanical nature. This might be due to a greater emphasis upon abstract problem solving in the training of electrical engineers. This result points to the importance of providing training in abstract, structured ways of problem solving (Kepner and Tregoe 1965).

If the present help facility were to be implemented in a real system, a number of additions could be made. For example, for each alarm the difference between the current and the normal value for the relevant system parameter might be listed in order to make it possible to more accurately estimate the seriousness of the deviation. In addition, the present help facility could be made more 'intelligent' if the order of the hypotheses that are listed would take into account not just the specific alarm for which help was requested but also the other alarms that are present. This might be done in a simple way by listing hypotheses that are mentioned in more than one of the help texts for the current alarms at the top.

Acknowledgements

The research in this article is partly based on a Master's Thesis by the second author submitted to Delft University of Technology, The Netherlands. The authors gratefully acknowledge the assistance of Rob Voorkamp and Kees Houttuin in the preparation of the experiments and the data collection. Jan Maarten Schraagen and Herke Schuffel provided helpful comments on a draft of this article.

Requests for reprints and correspondence concerning this article should be sent to Jeroen G. W. Raaijmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands.

References

BOER, J. P. A. 1992, Fault diagnosis in the ship's control centre of the M-frigate; a simulator study. [Foutdiagnose in de technische centrale van het M-fregat: een simulatorstudie]. Report IZF 1992 A-2, TNO Institute for Perception, Soesterberg, The Netherlands (in Dutch).

ELSHOUT, J. J. and VEENMAN, M. V. J. 1992, Relation between intellectual ability and working method as predictors of learning, *Journal of Educational Research*, **85**, 134–143.

KEPNER, C. H. and TREGOE, B. B. 1965, *The Rational Manager* (McGraw-Hill, New York).

MORRIS, N. M. and ROUSE, W. B. 1985a, Review and evaluation of empirical research in troubleshooting, *Human Factors*, **27**, 503–530.

MORRIS, N. M. and ROUSE, W. B. 1985b, The effects of type of knowledge upon human problem solving in a process control task, *IEEE Transaction on Systems, Man, and Cybernetics*, **15**, 698–707.

PATRICK, J. and HAINES, B. 1988, Training and transfer of fault-finding skill, *Ergonomics*, **31**, 193–210.

RASMUSSEN, J. 1981, Models of mental strategies in process plant diagnosis, in J. Rasmussen and W. B. Rouse (eds), *Human Detection and Diagnosis of System Failures* (Plenum Press, New York), 241–258.

RASMUSSEN, J. 1983, Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models, *IEEE Transactions on Systems, Man, and Cybernetics*, **13**, 257–266.

RASMUSSEN, J. 1986, *Information Processing and Human-Machine Interaction* (North-Holland, New York).

SCHAUBLE, L., GLASER, R., RAGHAVAN, K. and REINER, M. 1992, The integration of knowledge and experimentation in strategies in understanding a physical system, *Applied Cognitive Psychology*, **6**, 321–343.

VEENMAN, M. V. J. 1993, Intellectual ability and metacognitive skill: Determinants of discovery learning in computerized learning environments. Unpublished Ph.D. Thesis, University of Amsterdam, Amsterdam, The Netherlands.

VEENMAN, M. V. J. and ELSHOUT, J. J. 1991, Intellectual ability and working method as predictors of novice learning, *Learning and Instruction*, **1**, 303–317.

VERDUYN, W. W. 1993, Diagnostic aiding for the duty officer in the Ships Control Centre of the M-frigate [Diagnoseondersteuning voor de chef van de wacht in de technische centrale van het M-fregat]. Unpublished Master's Thesis, Delft University of Technology, Delft, The Netherlands (in Dutch).

YOON, W. C. and HAMMER, J. M. 1988, Deep-reasoning fault diagnosis: an aid and a model, *IEEE Transactions on Systems, Man, and Cybernetics*, **18**, 659–676.