Research Note 86-33

HUMAN PROBLEM SOLVING IN FAULT DIAGNOSIS TASKS

William B. Rouse and Ruston M. Hunt

Center for Man-Machine Systems Research
GEORGIA INSTITUTE OF TECHNOLOGY

Contracting Officer's Representative
Michael Drillings

BASIC RESEARCH
Milton S. Katz, Director

U. S. Army

Research Institute for the Behavioral and Social Sciences

April 1986

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the
Deputy Chief of Staff for Personnel

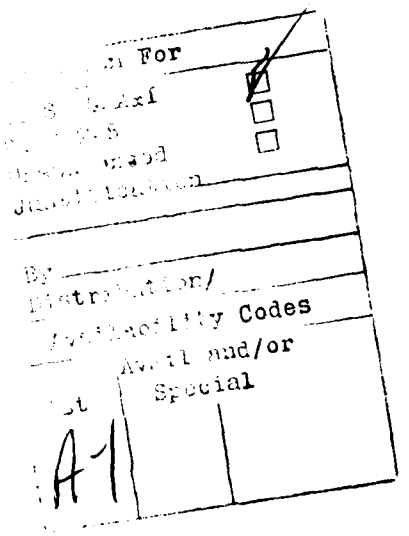
EDGAR M. JOHNSON
Technical Director

WM. DARRYL HENDERSON
COL, IN
Commanding

Research accomplished under contract for
the Department of the Army

Center for Man-Machine Systems Research,
Georgia Institute of Technology

Technical review by
Steve Kronheim



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARI Research Note 86-33	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) HUMAN PROBLEM SOLVING IN FAULT DIAGNOSIS TASKS		5. TYPE OF REPORT & PERIOD COVERED Final Report July 1979 - July 1982
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William B. Rouse and Ruston M. Hunt		8. CONTRACT OR GRANT NUMBER(s) MDA 903-79-C-0421
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Man-Machine Systems Research Georgia Institute of Technology Atlanta, Georgia 30332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q161102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS Army Research Institute for the Behavioral and Social Sciences. 5001 Eisenhower Avenue Alexandria, Virginia 22333-5600		12. REPORT DATE April 1986 13. NUMBER OF PAGES 54
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Contracting Officer's Representative, Michael Drillings.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Problem Solving
Fault Diagnosis
Troubleshooting,
Fuzzy Set Model,
Complexity Model,
Rule-Based Model,
Training.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report evaluates the nature of human problem solving abilities specifically related to fault diagnosis situations. Three types of fault diagnosis were analysed involving diagnosis utilizing both real equipment and computer simulated equipment failures. In addition, the investigators experimented with computer generated problem solving aids to supplement human decision making capacities in diagnostic tasks. Results of the intial investigation indicate that human problem solving tends to be highly

(con't)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

item #20 Abstract

context-specific but that pattern recognition capabilities are exceptional allowing for a high degree of accuracy in ambiguous problem solving situations. Both structured-oriented and strategy-oriented problem solving aids were analysed. Structured-oriented bookkeeping aids clearly improved performance, while strategy-oriented aids actually had a negative effect on transfer of training. This research effort is clearly relevant to military interests in effective training of optimal decision making in sub-optimal conditions. The results strongly support further research into the nature and training of effective problem solving. Keywords: → LD19

TABLE OF CONTENTS

INTRODUCTION

One of the reasons often given for employing humans in systems is their supposed abilities to react appropriately and flexibly in failure situations. On the other hand, one seems to hear increasingly about failure situations being aggravated by "human error". The apparent inconsistency of these two observations can cause one to wonder what role the human should actually play [Rasmussen and Rouse, 1981].

This question has led the authors and their colleagues to the pursuit of a series of investigations of human problem solving performance in fault diagnosis tasks. Using three different fault diagnosis scenarios, several hundred subjects (mostly maintenance trainees) have been studied in the process of solving many thousands of problems. The results of these studies have motivated the development of several mathematical models of human problem solving behavior. The three tasks, results of ten experiments, and five models are reviewed in this report.

Besides trying to assess problem solving abilities, considerable effort has also been invested in studying alternative methods of training humans to perform fault diagnosis tasks. One issue that has been particularly intriguing concerns the extent to which humans can be trained to have general, context-free problem solving skills. From a theoretical point of view, it is of fundamental interest to know whether skills are context-free or context-specific. From a practical perspective, this issue is perhaps even more important in terms of training

personnel to serve in multiple domains (e.g., to diagnose faults in a wide variety of systems). This report considers the extent to which the studies discussed here have provided an answer to the context-free versus context-specific question.

The overall goal of this research has been to determine an appropriate role for humans in failure situations and, to develop methods of training humans to fill that role. In a final section of this report, the variety of results presented here will be used as a basis for proposing how these issues should be resolved.

FAULT DIAGNOSIS TASKS

Three types of fault diagnosis task were used in this research. Two types involve computer simulations of network representations of systems in which subjects are required to find faulty components. The third type involves troubleshooting of real equipment. The three types of task represent a progression from a fairly abstract simulation that includes only one or two basic operations, to a somewhat realistic simulation and, finally, to real equipment.

TASK

In considering alternative fault diagnosis tasks for initial studies, one particular task feature seemed to be especially important. This feature is best explained with an example. When trying to determine why component, assembly, or subsystem A is producing unacceptable outputs, one may note that acceptable performance of A requires that components B, C, and D be performing acceptably since component A depends upon them. Further, B may depend on E, F, G, and H while C may depend on F and G, and so on. Fault diagnosis in situations such as this example involves dealing with a network of dependencies among components in terms of their abilities to produce acceptable outputs. The class of tasks described in this paragraph was the basis for the task chosen for initial investigations. Because this type of task emphasizes the structural properties of systems (i.e., relationships among components), the acronym chosen was TASK which stands for Troubleshooting by Application of

Structural Knowledge.

TASK involves fault diagnosis of graphically displayed networks. An example of TASK 1 is shown in Figure 1. These networks operate as follows. Each component has a random number of inputs. Similarly, a random number of outputs emanate from each component. Components are devices that produce either a 1 or Ø. An output of 1 denotes an acceptable output; Ø an unacceptable output. All outputs emanating from a component carry the value produced by that component.

A component will produce a 1 if: 1) All inputs to the component carry values of 1 and, 2) The component has not failed. If either of these two conditions are not satisfied, the components will produce a Ø. Thus, components are like AND gates. If a component fails, it will produce values of Ø on all the outputs emanating from it. Any components that are reached by these outputs will in turn produce values of Ø. This process continues and the effects of a failure are thereby propagated throughout the network.

A problem begins with the display of a network with the outputs indicated, as shown on the righthand side of Figure 1. Based on this evidence, the subject's task is to "test" connections between components until the failed component is found. The upper lefthand side of Figure 1 illustrates the manner in which connections are tested. An * is displayed to indicate that subjects can choose a connection to test. They enter commands of the form "component 1, component 2" and are

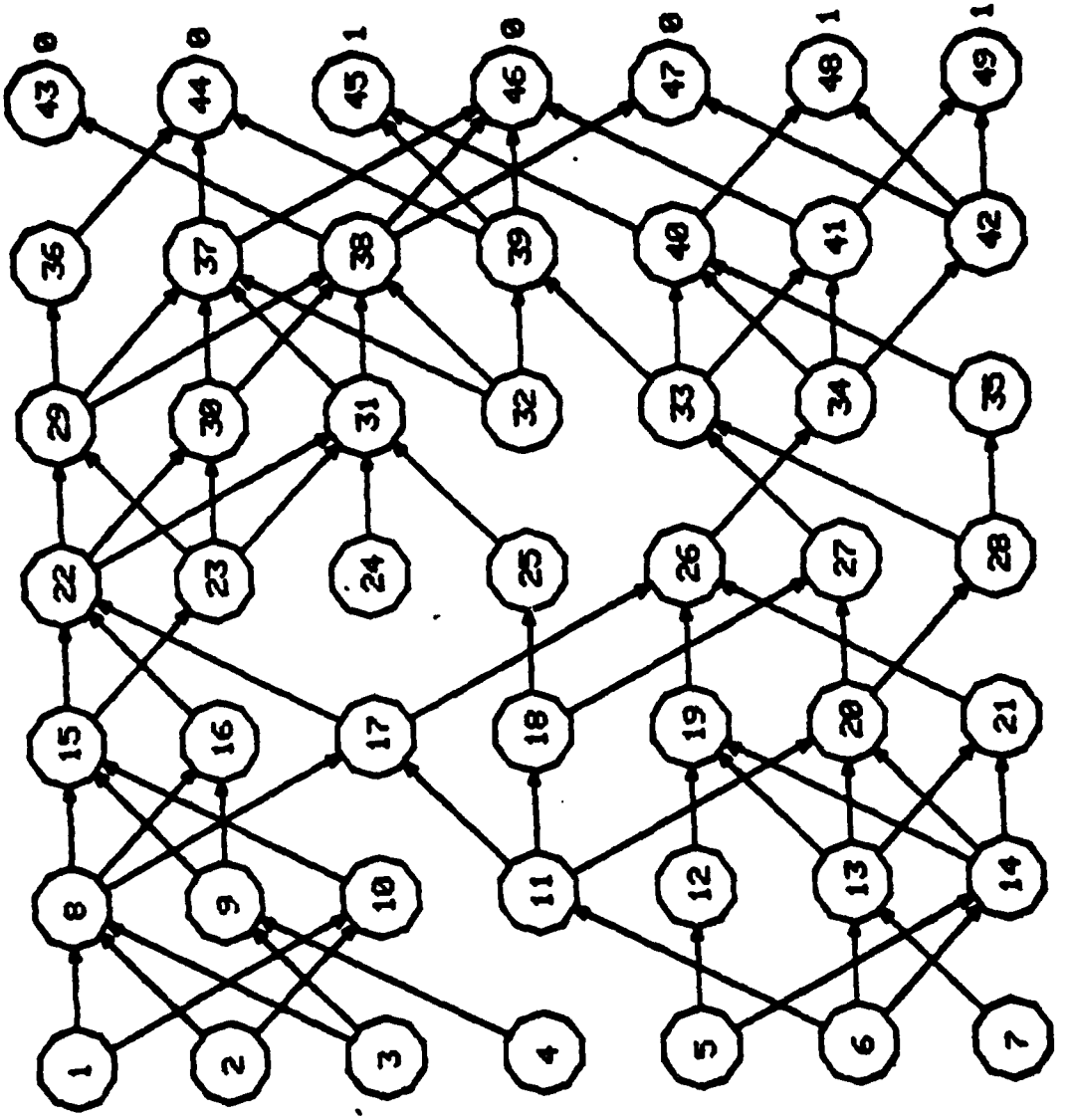


Figure 1. An Example of TASK 1

```
22, 30 = 1
23, 30 = 1
30, 38 = 0
31, 38 = 1
24, 31 = 1
25, 31 = 1
******** FAILURE ? 31

RIGHT!
```

then shown the value carried by the connection. If they respond to the * with a simple "return", they are asked to designate the failed component. Then, they are given feedback about the correctness of their choice (1). And then, the next randomly-generated problem (i.e., totally new) is displayed.

In the experiments conducted using TASK 1, computer aiding was one of the experimental variables. The aiding algorithm is discussed in detail elsewhere [Rouse, 1978a]. Succinctly, the computer aid is a somewhat sophisticated bookkeeper that uses the structure of the network (i.e., its topology) and known outputs to eliminate components that cannot possibly be the fault (i.e., by crossing them off). Also, it iteratively uses the results of tests (chosen by the subject) to further eliminate components from future consideration by crossing them off. In this way, the "active" network iteratively becomes smaller and smaller.

TASK 1 is fairly limited in that only one type of component is considered. Further, all connections are feed-forward and thus, there are no feedback loops. To overcome these limitations, a second version of TASK was devised.

Figure 2 illustrates the type of task of interest. This task is somewhat similar to TASK 1 in terms of using an acceptable/unacceptable dichotomy, requiring similar commands from subjects, and so on. Only the differences between TASK 1

(1) In the earlier experiments, subjects were not allowed to continue if their choice was incorrect; in the later experiments, they were instructed to continue until the failure was found.

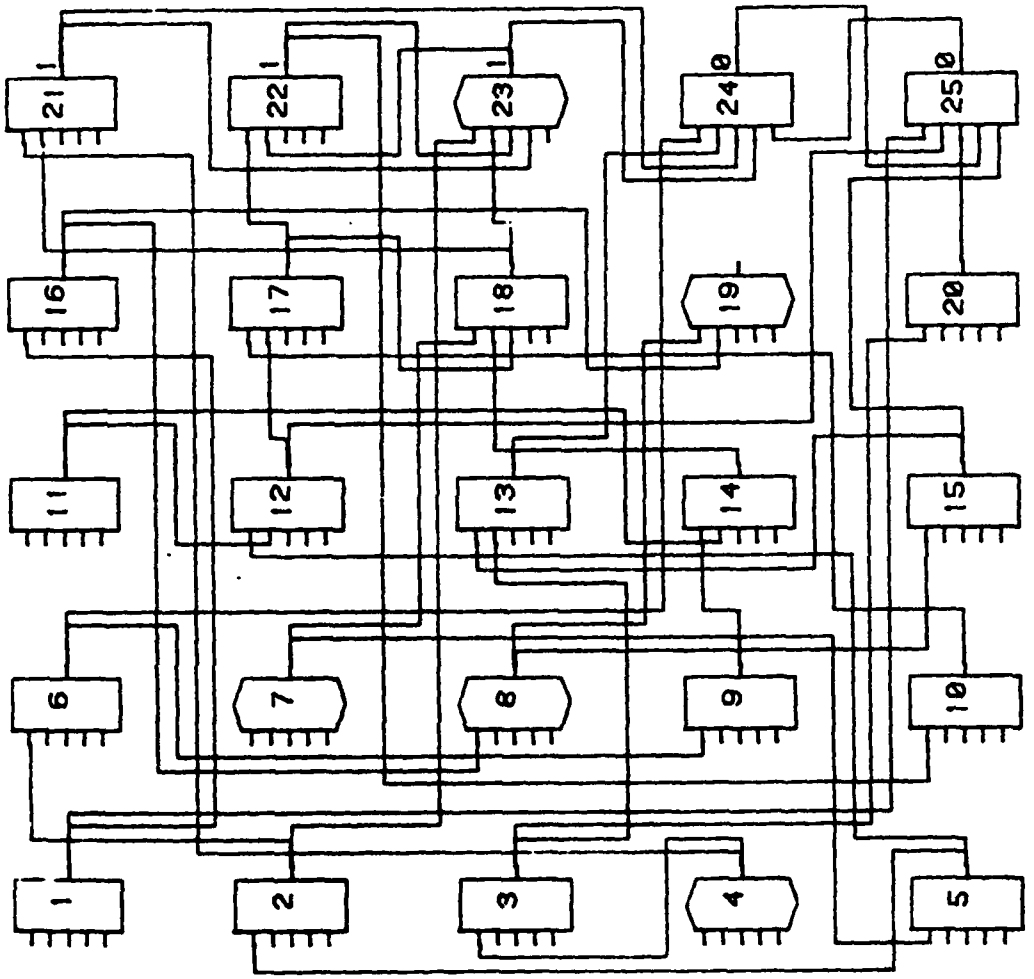


Figure 2. An Example of TASK 2

```
* 20 25 = 1
* * 13 24 = 0
* * * 15 13 = 0
* * * * 8 15 = 0
* * * * * 1 25 = 0
* * * * * FAILURE ? 1

RIGHT!
```

and TASK 2 are explained here.

A square component will produce a 1 if: 1) All inputs to the component carry values of 1 and, 2) The component has not failed. Thus, square components are like AND gates. A hexagonal component will produce 1 if: 1) Any input to the component carries a value of 1, and 2) The component has not failed. Thus, hexagonal components are like OR gates. For both AND and OR components, if either of the two conditions is not satisfied, the component will produce a 0.

The overall problem is generated by randomly connecting components. Connections to components with higher numbers (i.e., feed-forward) are equally likely with a total probability of p. Similarly, connections to components with lower numbers (i.e., feedback) are equally likely with a total probability of 1-p. The ratio p/(1-p), which is an index of the level of feedback, was one of the independent variables in the experiments to be discussed later. OR components are randomly placed. The effect of the ratio of the number of OR to AND components was also an independent variable in the experiments.

FAULT

TASK 1 and TASK 2 are context-free fault diagnosis tasks in that they have no association with a particular system or piece of equipment. Further, subjects never see the same problem twice. Thus, they cannot develop skills particular to one problem. Therefore, one must conclude that any skills that subjects develop have to be general, context-free skills.

However, real-life tasks are not context-free. And thus, one would like to know if context-free skills are of any use in context-specific tasks. In considering this issue, one might first ask: Why not train the human for the task he is to perform? This approach is probably acceptable if the human will in fact only perform the task for which he is trained. However, with technology changing rapidly, an individual is quite likely to encounter many different fault diagnosis situations during his career. If one adopts the context-specific approach to training, then the human has to be substantially retrained every time he changes situations.

An alternative approach is to train humans to have general skills which they can transfer to a variety of situations. Of course, they still will have to learn the particulars of each new situation, but they will not do this by rote. Instead, they will use this context-specific information to augment their general fault diagnosis abilities.

The question of interest, then, is whether or not one can train subjects to have general skills that are in fact transferrable to context-specific tasks. With the goal of answering this question in mind, another fault diagnosis task was designed [Hunt, 1979; Hunt and Rouse, 1981]. The acronym chosen for this task was FAULT which stands for Framework for Aiding the Understanding of Logical Troubleshooting.

Since FAULT is context-specific, one can employ hardcopy schematics rather than generating random networks online such as used with TASK. A typical schematic is shown in Figure 3. The subject interacts with this system using the display shown in Figure 4. The software for generating this display is rather general and particular systems of interest are completely specified by data files, rather than by changes in the software itself. Thus far, various automobile, aircraft, and marine systems have been simulated.

FAULT operates as follows. At the start of each problem, subjects are given rather general symptoms (e.g., will not light off). They can then gather information by checking gauges, asking for definitions of the functions of specific components, making observations (e.g., continuity checks), or by removing components from the system for bench tests. They also can replace components in an effort to make the system operational again.

Associated with each component are costs for observations, bench tests, and replacements as well as the a priori probability of failure. Subjects obtain this data by requesting information about specific components. The time to perform observations and tests are converted to dollars and combined with replacement costs to yield a single performance measure of cost. Subjects are instructed to find failures so as to minimize cost.

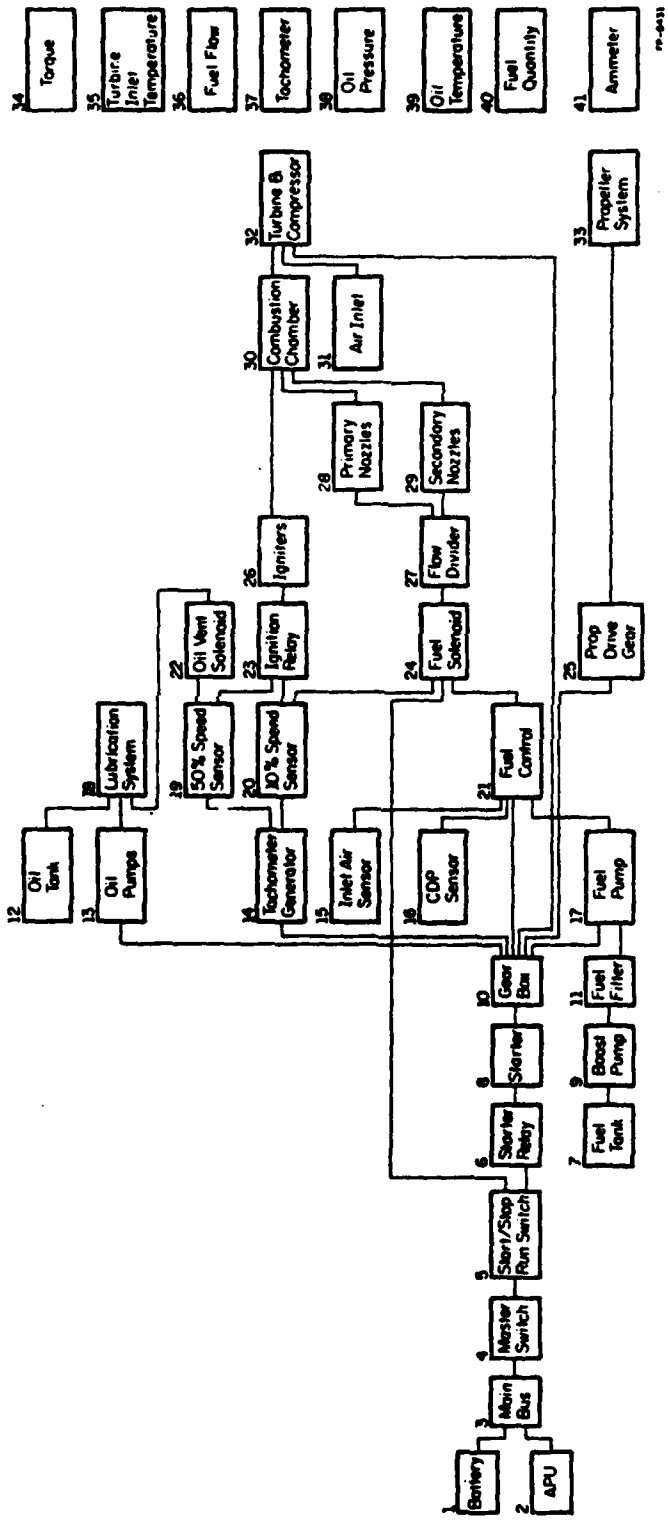


Figure 3. An Example FAULT System

System: Turboprop Symptom: Will not light off

You have six choices:

1	Observation	OX,Y
2	Information	IX
3	Replace a part	RX
4	Gauge reading	GX
5	Bench test	BX
6	Comparison	CX,Y,Z

(X,Y and Z are part numbers)

34	Torque	
35	Turbine Inlet Temp	Low
36	Fuel Flow	Low
37	Tachometer	Low
38	Oil Pressure	Normal
39	Oil Temperature	Normal
40	Fuel Quantity	
41	Ammeter	Normal

Your choice ...

Actions	Costs	Actions	Costs	Parts Replaced	Costs
4, 5 Normal	$ 1			14 Tach Generator	$ 199
26,30 Abnormal	$ 1				
14,20 Not aval	$ 0				
14 is Abnormal	$ 27				

FP-6441

Figure 4. The Fault Display

As with TASK, computer aiding was an independent variable in one of the experiments with FAULT [Hunt, 1981; Hunt and Rouse, 1982b]. The aiding scheme monitors subjects for inferential errors (i.e., seeking information that, by structural inference, is already available) and provides context-specific feedback concerning how the appropriate inference could be made. Aided subjects were also allowed to test the validity of hypotheses by asking the computer whether or not a particular component was in the feasible set of possible failures given the information collected up to that point.

Real Equipment

The experiments involving real equipment required subjects to diagnose failures in four and six cylinder engines typical of those used in modern general aviation aircraft [Johnson, 1980; Johnson and Rouse, 1982b]. The five problems chosen for study represented four engine subsystems: electrical, ignition, lubrication, and fuel. More specifically, the five problems studied were: 1) an open starter lead, 2) a defective spark plug wire, 3) an obstructed oil fitting, 4) a defective spark plug, and 5) an obstructed fuel line.

Subjects were required to observe malfunctioning (but operating) engines and, by appropriately choosing tests, identify the source of the problem. They were supplied with all of the tools and test equipment necessary to diagnose any fault that they might encounter. Technical manuals and related information were also available.

MEASURES OF PERFORMANCE

In the series of experiments to be discussed in the next section, the subjects' instructions varied as the series progressed. While the initial experiment emphasized minimizing the number of tests to diagnose the failure correctly, later experiments stressed minimum time and cost. All three of these measures reflect the product of fault diagnosis. While such measures may appropriately gauge the overall goals of fault diagnosis, product measures do not provide much insight into the process of fault diagnosis [Duncan and Gray, 1975; Brooke, et al., 1980]. Much finer-grained process measures are needed to provide the desired insights into human behavior. In this section, the way in which this issue was addressed is reviewed.

Dimensions of Performance

Analyses of the results of the initial experiments with TASK were limited to product measures, typically adjusted for problem difficulty by normalizing with respect to optimal performance. These measures appeared to be satisfactory until experiments with FAULT were conducted. It was then found that the product measures were much too sensitive to individual differences among problems and subjects.

This realization led to the development of a variety of fine-grained process measures [Hunt and Rouse, 1981]. One pair of these measures considers diagnostic costs greater than optimal (the minimum) and partitions these suboptimal costs into two

categories: errors and inefficiency. Errors are defined as actions that do not reduce the size of the feasible set of failures (i.e., non-productive actions). Inefficient actions are productive but not as productive as possible. Another fine-grained measure is the expected (as opposed to actual) information gain (in bits) per action. A third measure reflects subjects' allocation of expenditures among the types of action available.

The usefulness of these process measures motivated a comprehensive investigation of performance measures [Henneman, 1981; Henneman and Rouse, 1982]. A throrough review of the literature, as well as consideration of previous experience with TASK and FAULT, produced a set of twenty candidate measures. These measures were evaluated using data from two of the later experiments. Correlation, regression, and factor analyses were employed.

The results were unequivocal. Among the twenty measures, there are only three unique dimensions: time, errors, and inefficiency. Thus, a single product measure such as time or cost does not adequately describe human performance. This result also showed that the choice of process measures of errors and inefficiency, as well as the product measure of time, for the earlier studies with FAULT was very appropriate.

The emergence of errors as one of the primary dimensions of diagnostic performance led to three studies of human error and the development of a general methodology for analysis and

classification of human error [van Eekhout and Rouse, 1981; Johnson and Rouse, 1982a; Rouse and Rouse, 1982b]. An early version of this methodology was used to analyze the results of the first real equipment experiment and produced changes in the training methods that were subsequently shown to reduce substantially the frequencies of certain types of human errors. These results are reviewed in a later section.

Predictive Measures

It is interesting to consider the extent to which problem solving performance is correlated with a priori characteristics of subjects rather than the effects of training. To explore this issue, the performance of subjects on TASK, FAULT, and real equipment was correlated with twelve measures of ability, aptitude, and cognitive style [Henneman, 1981; Henneman and Rouse, 1982]. Results of standard scholastic aptitude tests were used as ability measures. A mechanical reasoning test was employed to obtain a measure of aptitude. Two dimensions of cognitive style were considered: impulsivity-reflectivity (via Matching Familiar Figures Test [Kagan, 1965]) and field dependence-independence (via Embedded Figures Test [Witkin, et. al., 1971]).

Results indicated that cognitive style was a much better predictor (Pearson's r ~ Ø.5) of problem solving performance than were the measures of ability and aptitude. It should be noted, however, that the trainees whose data was employed in this analysis had to meet certain standards of ability and aptitude

(but not style) in order to be accepted into the training program which was studied. Thus, the fairest conclusion seems to be that cognitive style becomes dominant once minimum standards of ability and aptitude are met.

Detailed statistical analyses of the cognitive style results were performed by partitioning trainees into impulsive and reflective groups, as well as field dependent and independent groups, and using analyses of variance with dependent measures of time, errors, and inefficiency [Rouse and Rouse, 1982a]. The strongest conclusion to result from this analysis was that impulsives made significantly more errors. Several interesting comparisons with results published in the cognitive style literature were also found.

A further analysis of performance changes over time indicated that reflective field independents were the best problem solvers, although the superiority of field independents over field dependents tended to decrease as experience was gained [Hunt, et al., 1981]. One can conjecture that the pattern recognition abilities of field dependents required more time to adapt to new problem domains; however, they did eventually adapt. On the other hand, the effects of impulsivity were not compensated for with practice.

EXPERIMENTS

Using the three tasks and variety of performance measures described in the last two sections, ten experiments were performed involving over 300 subjects who solved over 24,000 fault diagnosis problems. Over 90% of the subjects were trainees in an FAA certificate program in aircraft powerplant maintenance. The remainder were students or former students in engineering. In this section, the statistically significant results of these experiments will be reviewed.

Experiments one through five focused on problem solving performance with TASK. Experiments six through eight considered the relationships between TASK and FAULT performance. Experiments nine and ten studied transfer of training from TASK and/or FAULT to real equipment.

Experiment One

The first experiment utilized TASK 1 and considered the effects of problem size, computer aiding, and training. Problem size was varied to include networks with 9, 25, and 49 components. Computer aiding was considered both in terms of its direct effect on task performance and in terms of its effect as a training device [Rouse, 1978a].

Eight subjects participated in this experiment. The experiment was self-paced. Subjects were instructed to find the fault in the minimum number of tests while also not using an excessive amount of time and avoiding all mistakes. A transfer

of training design was used where one-half of the subjects were trained with computer aiding and then transitioned to the unaided task, while the other one-half of the subjects were trained without computer aiding and then transitioned to the aided task.

Results indicated that human performance, in terms of average number of tests until correct solution, deviated from optimality as problem size increased. However, subjects performed much better than a "brute force" strategy which simply traces back from an arbitrarily selected Ø output. This result can be interpreted as meaning that subjects used the topology of the network (i.e., structural knowledge) to a great extent as well as knowledge of network outputs (i.e., state knowledge).

Considering the effects of computer aiding, it was found that aiding always produced a lower average number of tests. However, this effect was not statistically significant. Computer aiding did produce a statistically significant effect in terms of a positive transfer of training from aided to unaided displays for percent correct. Specifically, percent correct was greater with aided displays (98% vs. 89%) and subjects who transferred aided-to-unaided were able to maintain the level of performance achieved with aiding.

Experiment Two

This experiment utilized TASK 1 and was designed to study the effects of forced-pacing [Rouse, 1978a]. Since many of the interesting results of the first experiment were most pronounced for large problems (i.e., those with 49 components), the second

experiment considered only these large problems. Replacing problem size as an independent variable was time allowed per problem, which was varied to include values of 30, 60, and 90 seconds. The choice of these values was motivated by the results of the first experiment which indicated that it would be difficult to solve problems in less than 30 seconds consistently while it would be relatively easy to solve problems in less than 90 seconds.

This variable was integrated into the experimental scenario by adding a clock to the display. Subjects were allowed one revolution of the clock in which to solve the problem. The circumference of the clock was randomly chosen from the three values noted above. If subjects had not solved the problem by the end of the allowed time period, the display was erased and they were asked to designate the failed component.

As in the first experiment, computer aiding and training were also independent variables. Twelve subjects participated in this experiment. Their instructions were to solve the problems within the time constraints while avoiding all mistakes.

Results of this experiment indicated that the time allowed per problem and computer aiding had significant effects on human performance. A particularly interesting result was that forced-paced subjects utilized strategies requiring many more tests than necessary (i.e., greater than self-paced subjects by a factor of approximately four). It appears that one of the effects of forced-pacing was that subjects chose to employ less

structural information in their solution strategies, as compared to self-paced subjects. While computer aiding resulted in significantly fewer tests (0.99 vs. 3.33) and a greater percent correct (89% vs. 80%), there was no positive (or negative) transfer of training for forced-paced subjects, indicating that subjects may have to be allowed to reflect on what computer aiding is doing for them if they are to gain transferrable skills. In other words, time pressure can prevent subjects from studying the task sufficiently to gain skills via computer aiding.

Experiment Three

Experiments one and two utilized students or former students in engineering as subjects. To determine if the results obtained were specific to that population, a third experiment investigated the fault diagnosis abilities of forty trainees in the fourth semester of a two-year FAA certificate program in aircraft powerplant maintenance [Rouse, 1979a].

The design of this experiment was similar to that of the first experiment in that TASK 1 was utilized and problem size, computer aiding, and training were the independent variables. However, only transfer in the aided-to-unaided direction was considered. Further, subjects' instructions differed somewhat in that they were told to find the failure in the least amount of time possible, while avoiding all mistakes and not making an excessive number of tests.

As in the first experiment, performance in terms of average number of tests until correct solution deviated from optimality as problem size increased. Computer aiding significantly decreased this deviation (0.60 vs. 1.71, or 65% better). Considering transfer of training, it was found that aided subjects utilized fewer tests to solve problems without computer aiding, particularly for the larger problems (1.11 vs. 2.12 tests greater than optimal). A very specific explanation of this phenomenon will be offered in a later discussion.

Experiment Four

Experiment four considered subjects' performance in TASK 2 [Rouse, 1979b]. Since the main purpose of this experiment was to investigate the suitability of a model of human decision making in fault diagnosis tasks that include feedback and redundancy, only four highly trained-subjects were used. The two independent variables included the aforementioned level of feedback (i.e., p/(1-p)) and the ratio of number of OR to AND components in a network of twenty-five components.

The results of this experiment indicated that increased redundancy (i.e., more OR components) significantly decreased the average number of tests (3.47 vs. 4.91) and average time until correct solution (63.3 sec vs. 101.7 sec) of fault diagnosis problems. While there were visible trends in performance as a function of the level of feedback, this effect was not significant. The reason for this lack of significance was quite clear. Two subjects developed a strategy that carefully

considered feedback while the other two subjects developed a strategy that discounted the effects of feedback. Thus, the average across all subjects was insensitive to feedback levels. One of the models to be described later yields a fairly succinct explanation of this result.

Experiment Five

The purpose of this experiment was to investigate the performance of maintenance trainees in TASK 2, while also trying to replicate the results of experiment three. Forty-eight trainees in the first semester of the previously noted FAA certificate program served as subjects [Rouse, 1979c].

The design involved a concatenation of experiments three and four. Thus, the experiment included two sessions. The first session was primarily for training subjects to perform the simpler TASK 1. Further, the results of the first session, when compared with the results of experiment three, allowed a direct comparison between first and fourth semester trainees.

The second session involved a between-subjects factorial design in which level of feedback and proportion of OR components were the independent variables. Further, training on TASK 1 (i.e., unaided or aided) was also an independent variable. Thus, the results of this experiment allowed assessment of transfer of training between two somewhat different tasks.

As in the previous experiments, TASK 1 performance in terms of average number of tests until correct solution deviated from optimality as problem size increased and, the deviation was substantially reduced with computer aiding (0.57 vs. 1.53, or 63% better). Computer aiding also resulted in faster solutions (46.5 sec vs. 62.1 sec). However, unlike the results from experiment three, there was no positive (or negative) transfer of training from the aided displays. This result as well as subjects' comments led to the conjecture that the first semester students perhaps differed from the fourth semester students in terms of intellectual maturity (i.e., the ability to ask why computer aiding was helping them rather than simply accepting the aid as a means of making the task easy).

On the other hand, TASK 2 provided some very interesting transfer of training results. In terms of average time until correct solution, subjects who received aiding during TASK 1 training were initially significantly slower in performing TASK 2. However, they eventually far surpassed those subjects who received unaided TASK 1 training. This initial negative transfer (13% slower) and then positive transfer (20% faster) is an interesting but puzzling phenomenon.

Experiment Six

This experiment considered subjects' abilities to transfer skills developed in the context-free TASK 1 and TASK 2 to the context-specific FAULT. Thirty-nine trainees in the fourth semester of the two-year FAA certificate program served as

subjects [Hunt, 1979; Hunt and Rouse, 1981].

The design of this experiment was very similar to previous experiments except the transfer trials involved FAULT rather than the context-free tasks. The FAULT scenarios used included an automobile engine and two aircraft powerplants, one of which was unfamiliar to trainees. Both TASK 1 and TASK 2 were used for the training trials. Overall, subjects participated in six sessions of ninety minutes in length over a period of six weeks.

As noted earlier, since initial analyses of the results indicated a very substantial degree of inter-subject and inter-problem variability, it was decided to employ more fine-grained measures for FAULT. One of these fine-grained measures involved partitioning subjects' suboptimality (i.e., expenditures greater than optimal) into those due to errors and those due to inefficiency. Another measure was the expected information gain (in bits) per action. A third measure reflected the subjects' allocation of expenditures among observations, bench tests, and unnecessary replacements.

Use of these fine-grained performance measures led to quite clear conclusions. Trainees who had received aided training with TASK 1 were consistently able to achieve significantly better performance on the powerplant problems ($513 vs. $578 for cost due to inefficiency), especially for problems involving less familiar powerplants. It was found that their suboptimality in terms of inefficiency could be attributed to their focusing on high cost, low information gain actions (i.e., bench tests and

replacements) to a much greater extent than the optimal solution.

Experiment Seven

The purpose of this experiment was to replicate experiment six using first semester rather than fourth semester maintenance trainees. Sixty trainees participated. The design of the experiment was very similar to experiment six except that only TASK 1 training was used. Further, one of the aircraft powerplant scenarios was changed to allow inclusion of a more sophisticated system [Hunt and Rouse, 1981].

The results for the first semester trainees were mixed with a substantial positive transfer of aided training in terms of inefficiency ($469 vs. $1266) and a slight negative transfer of training in terms of expected information gain (0.51 vs. 0.53 bits/action). However, as with the fourth semester trainees, inefficiency could be attributed to inappropriate choices of high cost, low information gain actions.

Experiment Eight

This experiment considered the effects of computer-aided training with FAULT. Thirty-four first semester maintenance trainees participated in ten problem solving sessions over a ten week period. Half of the subjects received aiding while the other half did not. The two groups were initially matched on the basis of TASK 1 performance. Problems on FAULT included six different automobile and aircraft systems, some of which were unfamiliar to subjects [Hunt, 1981; Hunt and Rouse, 1982b].

The results of this experiment indicated that aiding decreased suboptimality in terms of inefficient actions for both the familiar (4.20 vs. 4.73) and unfamiliar (4.47 vs. 4.90) systems. Aiding significantly reduced the frequency of errors for the unfamiliar systems (0.40 vs. 0.83). (It is important to note that the aiding was designed to reduce errors; benefits in terms of decreased inefficency were only a by-product of aiding.) Considering transfer from FAULT to TASK, subjects trained with aided FAULT had a lower frequency of errors with TASK (0.08 vs. 0.20).

Experiment Nine

The purpose of this experiment was to evaluate the transfer of training with TASK 1, TASK 2, and FAULT to real equipment [Johnson, 1980; Johnson and Rouse, 1982b]. Thirty-six fourth semester trainees participated as subjects. Each subject was allocated to one of the three training groups. Groups were balanced with respect to various a priori measures (e.g., grade point average). One group was trained using a sequence of TASK 1 and TASK 2 problems. Another group was trained with FAULT. The third group, the control group, received "traditional" instruction including reading assignments, video taped lectures, and quizzes. The transfer task involved the aforementioned five problems on two real aircraft engines.

Performance measures for the real equipment problems included an average performance index based on a fine-grained analysis of each action, overall adjusted cost (based on the

manufacturer's flat-rate manual), and an overall rating by an observer. Results indicated that traditional instruction was only superior if explicit demonstrations were provided for the exact failures to be encountered (i.e., three of the five real equipment problems). Otherwise, there were no significant differences among the three training methods.

More specifically, for the average performance index, which ranged from 1.0 to 5.0, the three problems which were explicitly demonstrated yielded 4.4 for traditional instruction and 3.8 for TASK and FAULT; the two problems that were not explicitly demonstrated yielded the non-significant difference of 4.4 for traditional instruction and 4.2 for TASK and FAULT. Thus, training with the computer simulations was as useful as traditional training as long as the latter form of instruction was general in nature (i.e., did not provide "cookbook" solutions for particular problems).

Experiment Ten

This experiment also considered transfer to real equipment, and compared a combination of TASK and FAULT to traditional instruction. Twenty-six fourth semester maintenance trainees served as subjects. One half of the subjects were trained with TASK/FAULT where FAULT was somewhat modified to include information on how tests are physically made and how results should be interpreted. The other half of the subjects received traditional instruction similar to that in experiment nine [Johnson and Rouse, 1982b].

Based on the same performance measures as used for experiment nine, it was found that the TASK/FAULT combination was equivalent to traditional instruction for all five problems, even those for which explicit solution sequences had been provided within the traditional instruction. More specifically, the average performance index was 4.2 for traditional instruction and 3.9 for TASK/FAULT, a difference which was not statistically significant. Thus, somewhat generalized training was found to be competitive with problem-specific training. The full implications of this result will be discussed in a later section.

MODELS OF HUMAN PROBLEM SOLVING

The numerous empirical results of the experimental studies discussed above are quite interesting and offer valuable insights into human fault diagnosis abilities. However, it would be more useful if one could succinctly generalize the results in terms of theories or models of human problem solving performance in fault diagnosis tasks (2). Such models will eventually be useful for predicting human performance in fault diagnosis tasks and, perhaps for evaluating alternative aiding schemes and training methods. More immediately, however, the models discussed here were of use for interpreting research results and defining the directions of the investigations.

Models of Complexity

It is interesting to consider why some fault diagnosis tasks take a long time to solve while others require much less time. Intuitively, it would seem to relate to problem complexity. This led to an investigation of alternative measures of complexity of fault diagnosis tasks [Rouse and Rouse, 1979].

A study of the literature of complexity led to the development of four candidate measures which were evaluated using the data from experiments three and five. It was found that two particular measures, one based on information theory and the

(2) For a review of the literature on models of human problem solving, especially for detection, diagnosis, and compensation for system failures, see Rouse [1982c].

other based on the number of relevant relationships within the problem, were reasonably good predictors (Pearson's r = Ø.84) of human performance in terms of time to solve TASK 1 and TASK 2 problems. The success of these measures appeared to be explained by the idea that they incorporated the human's understanding of the problem and specific solution strategy as well as the properties of the problem itself. Thus, complexity should be viewed as related to both the problem and problem solver.

Fuzzy Set Model

One can look at the task of fault diagnosis as involving two phases. First, given the set of symptoms, one has to partition the problem into two sets: a feasible set (those components which could be causing the symptoms) and an infeasible set (those components which could not possibly be causing the symptoms). Second, once this partitioning has been performed, one has to choose a member of the feasible set for testing. When one obtains the test result, then the problem is repartitioned, with the feasible set hopefully becoming smaller. This process of partitioning and testing continues until the fault has been localized and the problem is therefore solved.

If one views such a description of fault diagnosis from a purely technical point of view, then it is quite straightforward. Components either can or cannot be feasible solutions and the test choice can be made using some variation of the half-split technique. However, from a behavioral point of view, the process is not so clear cut.

Humans have considerable difficulty in making simple yes/no decisions about the feasibility of each component. If asked whether or not two components, which are distant from each other, can possibly affect each other, a human might prefer to respond "probably not" or "perhaps" or "maybe".

This inability to make strict partitions when solving complex problems can be represented using the theory of fuzzy sets [Rouse, 1980, 1982d]. Quite briefly, this theory allows one to define components as having membership grades between 0.0 and 1.0 in the various sets of interest. Then, one can employ logical operations such as intersection, union, and complement to perform the partitioning process. Membership functions can be used to assign membership grades as a function of some independent variable that relates components (e.g., "psychological distance"). Then, free parameters within the membership functions can be used to match the performance of the model and the human. The resulting parameters can then be used to develop behavioral interpretations of the results of various experimental manipulations.

Such a model was developed and compared to the results of experiments one, two, and four in terms of average number of tests to correctly diagnose faults in TASK 1 and TASK 2 [Rouse, 1978b, 1979b]. For TASK 1, the model and subjects differed by an average of only 5%. For TASK 2, with the exception of one trial where two of the subjects made many errors, the comparison was comparable.

Two particularly important conclusions were reached on the basis of this modeling effort. First, the benefit of computer aiding lies in its ability to make full use of l outputs shown in Figures 1 and 2, which humans tend to greatly under-utilize. Second, the different strategies of subjects in experiment four can be interpreted almost soley in terms of the ways in which they considered the importance of feedback loops.

It is useful to note here that these quite succinct conclusions, and others not discussed here [Rouse, 1978b, 1979b], were made possible by having the model parameters to interpret. The empirical results did not in themselves allow such tight conclusions.

Rule-Based Model

While the fuzzy set model has proven useful, one wonders if an even simpler explanation of human problem solving performance would not be satisfactory. With this goal in mind, a second type of model was developed [Pellegrino, 1979; Rouse, Rouse, and Pellegrino, 1980]. It is based on a fairly simple idea. Namely, it starts with the assumption that human problem solving involves the use of a set of situation-action rules (or heuristics) from which the human selects, using some type of priority or control structure [Newell and Simon, 1972; Waterman and Hayes-Roth, 1978; Rouse, 1980].

Based on the results of experiments three, five, and six, an ordered set of twelve rules was found that adequately describes TASK 1 performance, in the sense of making tests similar to those of subjects 89% of the time. Using a somewhat looser set of four rules, the match increases to 94%. For TASK 2, a set of five rules results in an 88% match. It was also found that the rank-ordering of the rules was affected by training, with aided training producing the more powerful rank-orderings.

The new insights provided by this model led to the development of a new notion of computer aided training. Namely, subjects were given immediate feedback about the quality of the rules which the model inferred they were using. They received this feedback after each test they made. Evaluation of this idea within experiment six resulted in the conclusion that rule-based aiding was counterproductive (36% more tests during training and 159% more upon transfer) because subjects tended to misinterpret the quality ratings their tests received. However, it appeared that ratings that indicated unnecessary or otherwise poor tests might be helpful. This hypothesis was tested and found to be true for FAULT in experiment eight.

Fuzzy Rule-Based Model

All of the modeling results noted above were based on problems involving TASK 1 and TASK 2. An attempt was made to apply these models, especially the rule-based model, to describe human performance using FAULT. Success was initially limited by what Rasmussen (1981) would call a shift from topographic to

symptomatic search strategies. In other words, once subjects shift from a context-free to context-specific situation, they attempt to use rules that map directly from the symptoms to the solution. In many cases, this mapping process can be adequately described by the earlier rule-based model. However, not infrequently it appears that subjects utilize what might be termed highly context-dominated rules, perhaps based on their experiences prior to training.

This dichotomy between symptomatic and topographic problem solving was formalized in a fuzzy rule-based model [Rouse and Hunt, 1981; Hunt, 1981; Hunt and Rouse, 1982a]. This model first attempts to find familiar patterns among the symptoms of the failure (i.e., among the state variables of the system). If a match is found, symptomatic rules (S-rules) are used to map directly from symptoms to hypothesized failure. If there are no familiar patterns among the state variables, the model uses topographic rules (T-rules) to search the structures (i.e., functional relationships) of the system. The rules chosen are those with highest membership in the fuzzy set of choosable rules which is defined as the intersection of fuzzy sets of recalled, applicable, useful, and simple rules.

This model was evaluated using the data from experiment eight. It was found that the model could exactly match approximately 50% of subjects' actions and utilize the same rules about 70% of the time. The evaluation of the model also provided a clear demonstration of subjects shifting from S-rules to T-rules when an unfamiliar system was encountered.

An Overall Model

All of the models discussed thus far were devised for the express purpose of providing direction to the studies with TASK, FAULT, and real equipment. Of course, considerable effort was also invested in attempting to generalize the model formulations. Thus, the fuzzy rule-based model, for example, certainly appears to be widely applicable. However, none of the models discussed earlier here can really be thought of as describing all of human problem solving.

The fifth and last model to be discussed here represents an attempt to synthesize a model capable of describing human problem solving in general [Rouse, 1982c]. This model is based on a thorough review of the problem solving literature and, to a great extent, the four earlier models. The model operates on three levels: 1) recognition and classification, 2) planning, and 3) execution and monitoring.

Recognition and classification is the process whereby humans determine the problem solving situations with which they are involved. Familiar situations may invoke a standard "frame" [Minsky, 1975] while unfamiliar situations may lead to the use of analogies or even basic principles of investigation. Planning may involve the use of familiar "scripts" [Schank and Abelson, 1979] or, if no script is available, require generation of alternatives, imagining of consequences, valuing of consequences, etc. [Johannsen and Rouse, 1979]. Execution and monitoring involves the S-rules and T-rules discussed earlier.

The model operates on the above three levels of problem solving by recursively using a single mechanism that is capable of recognizing both patterns of state information and patterns of structural information. By recursively and constantly accessing this single mechanism the model is capable of both hierachical [Sacerdoti, 1974] and heterarchical [Hayes-Roth and Hayes-Roth, 1979] problem solving. Simultaneous operation on multiple levels also allows the model to pursue multiple goals such as occur in dynamic systems where the problem solver must coordinate both diagnosing the source of the problem and keeping the system operating.

A particularly interesting aspect of this model's behavior, as well as that of humans, is its potential for making errors. The model has two inherent possibilities for causing errors. The first possibility relates to the model's recursive use of a single basic mechanism. As the model recursively invokes this mechanism, it needs a "stack" or some short-term memory for keeping track of where it is and how it got there. If short-term memory is limited, as it is in humans, the model may recurse its way into getting lost or, pursuing tangents from which it never returns. To constrain this phenomenon, it is more likely to forget one's umbrella than to forget to go to work.

The second possiblity for causing errors is the matching of irrelevant or inappropriate patterns. For example, the model, or a human, may be captured by an inappropriate but similar script or S-rule. As a result, the model may pursue an inappropriate path until it suddenly realizes, perhaps too late to be able to

recoup, that it has wandered far afield from where it thought it was headed.

The fact that this model has inherent possibilities for making errors, particularly somewhat subtle errors, provides an interesting avenue for evaluating the model. Most models are evaluated in terms of their abilities to achieve the same levels of desired task performance as humans. A much stronger test would involve determining if the model deviates from desired performance in the same way and for the same reasons as humans. The proposed model can potentially be evaluated in this manner.

However, this model has not yet been evaluated. Thus, at this point, it should mainly be viewed as a synthesis of the wide variety of experimental results and models reviewed here. However, considering the breadth of the investigations upon which it is based, including the extensive review of the literature, this model should also be viewed as much more than conjecture. Clearly, the next step should be evaluation of this model in a variety of problem solving domains.

DISCUSSION AND CONCLUSIONS

The overall results of this program of research roughly fall into three categories:

1. Results relating to human problem solving abilities

2. Concepts for training and aiding problem solvers

3. Implications for the role of humans in failure situations

In this final section of this report, the findings in these three areas will be reviewed.

Human Problem Solving Abilities

Humans are not optimal problem solvers, although they are rational and usually systematic. In general, their deviation from optimality is related to how well they understand the problem, rather than being solely related to properties of the problem. More specifically, suboptimality appears to be due to a lack of awareness (or inability when forced-paced) of the full implications of available information. For example, humans have a great deal of difficulty utilizing information about what has not failed in order to reduce the size of the feasible set.

Human problem solving tends to be context-dominated with familiar, or even marginally familiar, patterns of contextual cues prevailing in most problem solving. Humans can, however, successfully deal with unfamiliar problem solving situations, which is a clear indication that human problem solving skills

cannot be totally context-specific. Their degree of success with unfamiliar problems depends on their abilities to transition from state-oriented to structure-oriented problem solving. Humans' abilities in the latter mode are highly related to their rank-ordering of rules rather than simply the number of rules available.

Thus, humans' cognitive abilities for problem solving are definitely limited. However, humans are exquisite pattern recognizers and can cope reasonably well with ill-defined and ambiguous problem solving situations. These abilities are very important in many real life fault diagnosis tasks. What are needed, then, are methods for overcoming humans' cognitive limitations in order to be able to take advantage of humans' cognitive abilities.

Concepts for Training and Aiding

Throughout this program of research, a variety of schemes have emerged for helping humans to overcome the limitations summarized above. These schemes have been evaluated both as aids during problem solving and as training methods, with evaluation occurring upon transfer to situations without the aids. As noted in previous sections, three types of aid were developed and evaluated.

The first type of aid was implemented within TASK and uses the structure of the network to determine the full implications of the symptoms, as well as each test, with respect to reduction of the size of the feasible set. Basically, this aid is a

bookkeeper that does not utilize any information which subjects do not have; it just consistently takes full advantage of this information.

The second type of aid was also implemented within TASK. It evaluates each action by subjects, as they occur, and provides reinforcement in proportion to the degree to which the action is consistent with a context-free optimal strategy. For erroneous (i.e., non-productive) actions, subjects receive feedback that simply notes, but does not explain, their errors. For inefficient (i.e., productive but far from optimal) actions, subjects receive feedback denoting their choices as poor or fair. Optimal or near optimal actions yield feedback indicating the choices to be good or excellent.

The third type of aid was implemented in FAULT. This aid monitors subjects' actions and checks for context-free inferential errors (i.e., errors in the sense of not using the structure of the FAULT network to infer membership in the feasible set). While the aiding is context-free, it explains the nature of the error within the context of the problem (i.e., in terms of the structural implications of the previous actions taken). Thus, the feedback received by subjects not only indicates the occurrence of an error, but also includes a context-specific explanation of why an error has been detected.

The first and third types of aid can both be viewed as structure-oriented bookkeeping aids, while the second type of aid is more strategy-oriented. The results of evaluating these aids

were quite clear. The bookkeeping methods consistently improved performance, both while they were available and upon transfer to unaided problem solving. The strategy-oriented aid degraded performance and resulted in negative transfer of training, providing clear evidence of the hazards of only reinforcing optimal performance.

In studies involving transfer from aided TASK to unaided TASK, aided TASK to unaided FAULT, and aided FAULT to unaided FAULT and unaided TASK, positive transfer of training was usually found with the effects most pronounced for unfamiliar systems and fine-grained performance measures. Thus, the evidence is quite clear that humans can be trained to have context-free problem solving skills that, at least partially, help them to overcome the limitations discussed earlier in this section.

Considering transfer from TASK and/or FAULT to real equipment, the results show that training based on simulations such as TASK and FAULT are competitive with traditional instruction, even when traditional instruction provides explicit solution procedures for the failures to be encountered. However, the issue is not really one for TASK versus FAULT versus traditional instruction. The important question is how these training technologies should be combined to provide a "mixed-fidelity" training program that capitalizes on the advantages of each technology [Rouse, 1982b]. This mixed-fidelity approach can provide trainees with problem solving principles as well as procedures. Also, it can result in a re-ordering of rules and not just the acquisition of more rules.

Thus, this approach can also help humans to overcome the previously discussed limitations. Finally, the mixed-fidelity approach can lead to cost savings since a training program need not rely solely on high-fidelity training devices.

Somewhat as a by-product of this research, a considerable amount was learned about evaluation of training programs [Rouse, 1982a]. Perhaps surprisingly, most evaluation efforts in the past have limited consideration to whether or not trainees learn to use the training technology successfully. Few studies have focused on transfer out of the training environment, and even fewer have looked at long-term on-the-job performance. Two of the studies reported here concentrated on transfer to real equipment; a study currently being planned will emphasize on-the-job performance.

One of the key aspects of evaluation is the definition of performance measures. The series of studies reviewed in this report began with the use of rather global measures and evolved to the use of very fine-grained measures where, for example, human error was classified using six general and thirty-one specific categories [Rouse and Rouse, 1982b]. It appears that this detailed level of analysis is very necessary if inadequacies in training programs are to be identified and remedied.

Finally, it should be noted that the model-based approach adopted for these investigations appears to have been a crucial element in their success. The evolving set of models provided succinct interpretations of results and, consequently, generated

very crisp hypotheses which focused subsequent investigations. Further, the models contributed to building an overall conceptual view of human problem solving.

The Role of Humans in Failure Situations

Based on the foregoing review of tasks, performance measures, experiments, and models, it seems reasonable to conclude with a discussion of the implications of these results for defining the role of humans in failure situations. As noted in the Introduction, there appears to be a tradeoff between the benefits of humans' unique abilities and the cost of their limitations. Resolving this tradeoff is tantamount to defining the role of humans.

One approach to dealing with this issue is to attempt to automate all fault diagnosis. Unfortunately, what this leads to is automation of routine diagnostic tasks and the human having responsibility for the more difficult problems. As a result, humans perform diagnostic tasks much less frequently; however, when humans must perform the diagnosis, the problem is likely to be very difficult, perhaps even involving untangling of the results of abortive attempts of the computer to diagnose the failures. This is a clear violation of good human factors engineering design principles.

A more appropriate approach is to emphasize computer aiding rather than computerizing. Results reported here indicate that computers can aid humans during training in terms of enhancing general problem solving skills and, during diagnosis by

performing bookkeeping functions and monitoring actions to assure that choices are productive. This approach leads to a perspective of humans controlling the problem solving process with sophisticated computer systems providing assistance. As a result, system designers can take advantage of human abilities while avoiding the effects of human limitations.

ACKNOWLEDGEMENTS

This research was mainly supported by the U.S. Army Research Institute for the Behavioral and Social Sciences under Contract No. MDA-903-79-C-0421 (1979-1982) and Grant No. DAHC 19-78-G-0011 (1978-1979). Some of the earlier work reported here was supported by the National Aeronautics and Space Administration under Ames Grant No. NSG-2119 (1976-1978). The authors gratefully acknowledge the many contributions of their colleagues William B. Johnson, Sandra H. Rouse, Richard L. Henneman, and Susan J. Pellegrino.

REFERENCES

1. Brooke, J.B., Duncan, K.D. and Cooper, C. 1980. "Interactive Instruction in Solving Fault Finding Problems," International Journal of Man-Machine Studies, Vol. 12, pp. 217-227.

2. Duncan, K.D. and Gray, M.J. 1975. "An Evaluation of a Fault Finding Training Course for Refinery Operators," Journal of Occupational Psychology, Vol. 48, pp. 199-218.

3. Hayes-Roth, B. and Hayes-Roth, F. 1979. "A Cognitive Model of Planning," Cognitive Science, Vol. 3, pp. 275-310.

4. Henneman, R.L. 1981. Measures of Human Performance in Fault Diagnosis Tasks. Urbana, IL: University of Illinois at Urbana-Champaign, M.S.I.E. Thesis.

5. Henneman, R.L. and Rouse, W.B. 1982. "Measures of Human Performance in Fault Diagnosis Tasks," submitted for publication.

6. Hunt, R.M. 1979. A Study of Transfer of Training from Context-Free to Context-Specific Fault Diagnosis Tasks. Urbana, IL: University of Illinois at Urbana-Champaign, M.S.I.E. Thesis.

7. Hunt, R.M. 1981. Human Pattern Recognition and Information Seeking in Simulated Fault Diagnosis Tasks. Urbana, IL: University of Illinois at Urbana-Champaign, Ph.D. Thesis.

8. Hunt, R.M. and Rouse, W.B. 1981. "Problem Solving Skills of Maintenance Trainees in Diagnosing Faults in Simulated Powerplants," Human Factors, Vol. 23, No. 3, pp. 317-328.

9. Hunt, R.M. and Rouse, W.B. 1982a. "A Fuzzy Rule-Based Model of Human Problem Solving," Proceedings of the American Control Conference, Washington, June.

10. Hunt, R.M. and Rouse, W.B. 1982b. "Computer-Aided Fault Diagnosis Training," submitted for publication.

11. Hunt, R.M., Henneman, R.L. and Rouse, W.B. 1981. "Characterizing the Development of Human Expertise in Simulated Fault Diagnosis Tasks," Proceedings of the International Conference on Cybernetics and Society, Atlanta, October, pp. 369-374.

12. Johannsen, G. and Rouse, W.B. 1979. "Mathematical Concepts for Modeling Human Behavior in Complex Man-Machine Systems," Human Factors, Vol. 21, pp. 733-747.

13. Johnson, W.B. 1980. Computer Simulations in Fault Diagnosis Training: An Empirical Study of Learning Transfer from Simulation to Live System Performance. Urbana, IL: University of Illinois at Urbana-Champaign, Ph.D. Thesis.

14. Johnson, W.B. and Rouse, W.B. 1982a. "Analysis and Classification of Human Errors in Troubleshooting Live Aircraft Powerplants," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-12, No. 3.

15. Johnson, W.B. and Rouse, W.B. 1982b. "Training Maintenance Technicians for Troubleshooting: Two Experiments with Computer Simulations," Human Factors, Vol. 24, No. 3.

16. Kagan, J. 1965. "Individual Differences in the Resolution of Response Uncertainty," Journal of Personality and Social Psychology, Vol. 2, No. 2, pp. 154-160.

17. Minsky, M. 1975. "A Framework for Representing Knowledge." in P.H. Winston, Ed., The Psychology of Computer Vision. New York: McGraw-Hill.

18. Newell, A. and Simon, H.A. 1972. Human Problem Solving. Englewood Cliffs, NJ: Prentice-Hall.

19. Pellegrino, S.J. 1979. Modeling Test Sequences Chosen by Humans in Fault Diagnosis Tasks. Urbana, IL: University of Illinois at Urbana-Champaign, M.S.I.E. Thesis.

20. 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. New York: Plenum Press.

21. Rasmussen, J. and Rouse, W.B. Eds., 1981. Human Detection and Diagnosis of System Failures. New York: Plenum Press.

22. Rouse, S.H. and Rouse, W.B. 1982a. "Cognitive Style as a Correlate of Human Problem Solving Performance in Fault Diagnosis Tasks," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-12, No. 5.

23. Rouse, W.B. 1978a. "Human Problem Solving Performance in a Fault Diagnosis Tasks," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-8, No. 4, pp. 258-271.

24. Rouse, W.B. 1978b. "A Model of Human Decision Making in a Fault Diagnosis Task," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-8, No. 5, pp. 357-361.

25. Rouse, W.B. 1979a. "Problem Solving Performance of Maintenance Trainees in a Fault Diagnosis Task," Human Factors, Vol. 21, No. 2, pp. 195-203.

26. Rouse, W.B. 1979b. "A Model of Human Decision Making in Fault Diagnosis Tasks that Include Feedback and Redundancy," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-9, No. 4, pp. 237-241.

27. Rouse, W.B. 1979c. "Problem Solving Performance of First Semester Maintenance Trainees in Two Fault Diagnosis Tasks." Human Factors, Vol. 21, No. 5, pp. 611-618.

28. Rouse, W.B. 1980. Systems Engineering Models of Human-Machine Interaction. New York: North-Holland.

29. Rouse, W.B. 1982a. "Design, Implementation, and Evaluation of Approaches to Improving Maintenance Through Training," Proceedings of the Design for Maintainers Conference, Pensacola, March.

30. Rouse, W.B. 1982b. "A Mixed-Fidelity Approach to Technical Training," Journal of Educational Technology Systems, Vol. 11.

31. Rouse, W.B. 1982c. "Models of Human Problem Solving: Detection, Diagnosis, and Compensation for Systems Failures." Proceedings of the IFAC Conference on Analysis, Design, and Evaluation of Man-Machine Systems, Baden-Baden, F.R. Germany, September.

32. Rouse, W.B. 1982d. "Fuzzy Models of Human Problem Solving," in P.P. Wang, Ed., Advances in Fuzzy Set Theory and Applications. New York: Plenum Press.

33. Rouse, W.B. and Hunt, R.M. 1981. "A Fuzzy Rule-Based Model of Human Problem Solving in Fault Diagnosis Tasks." Proceedings of the Eighth Triennial World Congress of IFAC. Kyoto, Japan,

34. Rouse, W.B. and Rouse, S.H. 1979. "Measures of Complexity of Fault Diagnosis Tasks," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-9, No. 11, pp. 720-727. (also appears in B. Curtis, Ed., Human Factors in Software Development. Silver Spring, MD: IEEE Computer Society Press, 1981.)

35. Rouse, W.B. and Rouse, S.H. 1982b. "Analysis and Classification of Human Error," submitted for publication.

36. Rouse, W.B., Rouse, S.H. and Pellegrino, S.J. 1980. "A Rule-Based Model of Human Problem Solving Performance in Fault Diagnosis Tasks," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-10, No. 7, pp. 366-376, (also appears in B. Curtis, Ed., Human Factors in Software Development. Silver Spring, MD: IEEE Computer Society Press, 1981.)

37. Sacerdoti, E.D. 1974. "Planning in a Hierarchy of Abstraction Spaces," Artificial Intelligence. Vol. 5, pp. 115-135.

38. Schank, R.C. and Abelson R.P. 1977. Scripts, Plans, Goals, and Understanding. Hillsdale, NJ: Lawrence Erlbaum.

39. van Eekhout, J.M. and Rouse, W.B. 1981. "Human Errors in Detection, Diagnosis, and Compensation for Failures in the Engine Control Room of a Supertanker," IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-11, No. 12, pp. 813-816.

40. Waterman, D.A. and Hayes-Roth, F., Eds., 1978. Pattern-Directed Inference Systems. New York: Academic Press.

41. Witkin, H.A., Ottman, P.K., Raskin, E. and Karp, S.A. 1971. A Manual for the Embedded Figures Test. Palo Alto, CA: Consulting Psychologists Press, Inc.