questions, that is, the criteria it addresses are not the right ones; 2) the process addresses the appropriate concerns but fails to utilize information correctly; or 3) the data that would be necessary to recognize a hazard are not available. The first possibility argues for including in any process as many criteria as can be appropriately incorporated and analyzed. The second possibility points out the need for careful development of priority-setting mechanisms and particularly for careful development of any algorithms by which criteria might be aggregated.

The last possibility is, of course, real and quite important. However, it should be recognized that this is inherent in the nature of the problem and will plague any effort to set priorities; it is not a problem which is peculiar to mechanical priority-setting methods. Any data analysis system can only analyze data which are available. If the available data indicate, for example, only that a substance is made in small quantities at two sites, the most that any system could be expected to do is indicate that currently available information presents no cause for concern.

Of course, it is not unimportant that potentially hazardous chemicals may go undetected due to an initial lack of information. It is very important, and argues, for example, that the initial sorting step should not irrevocably assign chemicals to a low priority category. Low priority chemicals should always remain open to further consideration if more data or new techniques (e.g., quantitative structure/activity relationship methods which seem so promising) become available.

The most important implication of the high probability that some potentially hazardous substances will be missed by any priority-setting method, is the incentive it creates for increasing the minimum data set available for all commercial chemicals. Without some information on both the toxicity of a chemical, and the amounts of it to which people are exposed, no real measure of its potential hazard is possible. Information on both toxicity and exposure, sufficient for accurate priority-setting, is only occasionally available at present.

Conclusion

Screening lists of chemicals is a challenging and, in many ways, unique problem. Unfortunately, the several programs of the federal government which face this problem will not be able to wait for the development of a theoretical foundation on which to base their efforts. The statutory deadlines under which these agencies operate do not allow for this luxury. The pressure is to do something, anything, but to put a system together to set priorities. There are few past efforts and even fewer theoretical explorations of this type of problem which can serve as guidance when federal employees or their hired contractors set to work, with little or no relevant experience, to develop the first priority-setting process they will ever use. The opportunities are enormous for doing something illogical or not well thought out. Hopefully this problem will attract the attention of scholars well versed in decision theory, as they probably are best trained to develop the necessary paradigm. In addition, the expertise of chemists, toxicologists, and others will be necessary because of the highly technical nature of the considerations involved. Over the next several years, this should be an exciting field in which to work.

References

[1] Steven L. Wilhelm, "Selecting priorities from large sets of alternatives: The case of toxic substances regulation," Ph.D. dissertation, Brown Univ. Providence, RI, (1981). Available from University Microfilms.

[2] "The relevance of multiple criteria decision making to priorities for chemical regulation: An annotated bibliography," Steven L. Wilhelm Ed., prepared by Tracor-Jitco Inc. for the U.S. Environmental Protection Agency, EPA Report EPA 560/7-81-005, July 1981.

[3] The Toxic Substances Control Act, PL 94-469, 15 USC 2601 *et. seq.*, 1976.

[4] Marilyn C. Bracken *et al.*, "Chemical substances information network," vol. 1, MITRE Corp., Washington, DC, MTR-7558, June 1977.

[5] "Chemical selection working group, criteria and procedures for chemical selection," submitted to the Clearinghouse Chemical Selection Subgroup, National Cancer Institute, Bethesda, MD.

[6] Interagency Testing Committee, "Initial report to the administrator," U.S. Environmental Protection Agency, Washington, DC, 42 FR 55026, 1977.

[7] Robert H. Ross and Paul Lu, "Chemical scoring system development," rep., Oak Ridge National Lab. to the U.S. Environmental Protection Agency, interagency agreement no. 79-D-X0856, EPA TSCA Industry Assistance Office, June 1981.

[8] John N. Gevertz and Elaine Bild, "Chemical selection methods: An annotated bibliography," U.S. Environmental Protection Agency, Rep., EPA 560/TIIS-80-001, Nov. 1980.

[9] C. Jackson Grayson, *Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators.* Cambridge, MA: Harvard Business School, 1960.

[10] C. E. Gearing, T. Var, and W. W. Swart, "A decision structure for tourist investment calculations," *Tourist Review*, no. 1, 1972.

[11] W. W. Swart, C. E. Gearing, and T. Var, "A dynamic programming–integer programming algorithm for allocating touristic investments," *Tourist Review*, no. 2, 1972.

[12] William L. Wilke and Edgar A. Pessemier, "Issues in marketing's use of multi-attribute attitude models," *J. Marketing Res.*, vol. 10, pp. 428–441, 1973.

[13] Don Pekelman and Subruta K. Sen, "Mathematical programming models for the determination of attribute weights," *Management Science*, vol. 20, no. 8, p. 1217–1229, 1974.

[14] Steven L. Wilhelm, "The OTS exposure estimation methodology," in "Selecting Priorities from large sets of alternatives: The case of toxic substances regulation," Ph.D. dissertation, Brown Univ., Providence, RI, 1981, Appendix A.

Human Errors in Detection, Diagnosis, and Compensation for Failures in the Engine Control Room of a Supertanker

JOOST M. VAN EEKHOUT AND WILLIAM B. ROUSE,
SENIOR MEMBER, IEEE

Abstract—**Seven crews of professional engineering officers were studied performing the task of coping with failures in a high-fidelity supertanker engine control room simulator. Measurement methods included verbal protocols, computer logs of all discrete events, interviews, questionnaires, and observer ratings. The resulting data were analyzed for human errors which were classified using a scheme developed within this study. It was found that errors associated with inappropriate identification of the failure were highly correlated with a lack of knowledge of the functioning of the basic system as well as the automatic controllers within the system. Further, errors related to execution of procedures were highly correlated with inadequacies of the layout of the control panel and simulator fidelity inadequacies. Based on these results, it is concluded that operator training should have increased emphasis on the knowledge necessary for dealing with failure situations.**

Introduction

Increasingly one hears about failures of technical systems being caused or aggravated by "human error" [1]. While this trend might lead one to believe that humans are subject to random onsets of mistakes, it is also possible to view many human errors as systematically caused by factors related to system design and operator training. This correspondence discusses a study that supports the latter perspective.

Manuscript received March 10, 1981; revised August 10, 1981.

J. M. van Eekhout was with the Delft University of Technology, Delft, The Netherlands. He is now with the Royal Dutch Navy, Den Helder, The Netherlands.

W. B. Rouse was with the Delft University of Technology, The Netherlands, and the Mechanical and Industrial Engineering Department, University of Illinois, Urbana, IL. He is now with the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

0018-9472/81/1200-0813$00.75

A prerequisite to studying human error is a recognition that not all errors are the same in terms of cause, mechanism, and consequences. Many investigators have recognized this issue, and a variety of error classification schemes have resulted. While many of these schemes have considerable merit [2], the ideas of Rasmussen and his colleagues were especially useful in the context of interest in this correspondence [3], [4]. Rasmussen's classification system is oriented towards categorization of human errors in nuclear power plant operations in terms of the environment, staffing, particular situation, and specific events. Of more importance, it explicitly treats the causes, mechanisms, and consequences of human errors. The error classification system presented in this correspondence can be viewed as an extension of Rasmussen's scheme to the marine engineering domain.

The goal of this correspondence is to present a systematic study of the types of errors committed by professional marine engineering officers in the course of detecting, diagnosing, and compensating for failures in the engine control room of a supertanker. An error classification system suitable for this domain is presented and a methodology for identification and classification of errors discussed. Finally, several factors that may contribute to the occurrence of human errors are discussed and approaches to ameliorating their effects considered.

Method

The engine control room simulator employed for this study is at the TNO (Dutch Organization for Applied Scientific Research) Institute for Mechanical Constructions at Delft, The Netherlands. The simulator includes the key portions of a marine steam power plant and is modeled after the machinery of the "Esso Wilhelmshaven," a crude container carrier of 250 000 tons. The simulator consists of

1) a control room equipped with the main instrument console and electrical switchboards,
2) three panels in an adjacent room which represent machinery instruments and controls that can only be accessed by leaving the control room,
3) an instructor's cabin with the simulator control console,
4) a computer which employs a mathematical model of the simulated power plant to convert control inputs to instrument readings (i.e., outputs).

A functional block diagram of the steam power plant is shown in Fig. 1. Operator performance in dealing with the following six different failures in this system was studied.

1) Failure of the boiler drum level transmitter: This results in the feedwater valve (FWV) fully opening, which causes an increase in the water level in the boiler and leads to a "high boiler level" alarm and possibly a turbine trip. As a side effect of the FWV fully opening, water will be drawn from the deaerator, perhaps resulting in a "low deaerator level" alarm, and cause the makeup valve to open to feed the condensor and subsequently the deaerator. The appropriate operator action is to switch the FWV from automatic to manual and control the boiler level manually.

2) Failure of running fuel oil pump: This results in the loss of fuel pressure, tripping of the burners, and "low fuel oil pressure" alarm. Operator should stop the pump, inspect the engine room for leaks, purge the boiler of gases, start the backup pump, and restart the boiler.

3) Failure of automatic changeover when feedwater pump (FWP) fails: This results in FWV fully opening, decreased FWP discharge pressure, and decreased steam flow to FWP. Alarm for "low FWP discharge pressure" is activated and, if the failure is not quickly found, the "low boiler level" alarm is activated. Operator should manually start lube oil pump for backup FWP and then start FWP.

4) Failure of differential pressure (dp) transmitter: This results in steam-driven FWP speeding up to produce desired dp as well as increased steamchest pressure and FWP discharge pressure. Increased flow via FWP results in transient decrease of deaerator level for which makeup valve will open to compensate. Alarms for "high boiler level" and "low deaerator level" may be activated. Operator should switch the FWP from automatic to manual and control it manually.

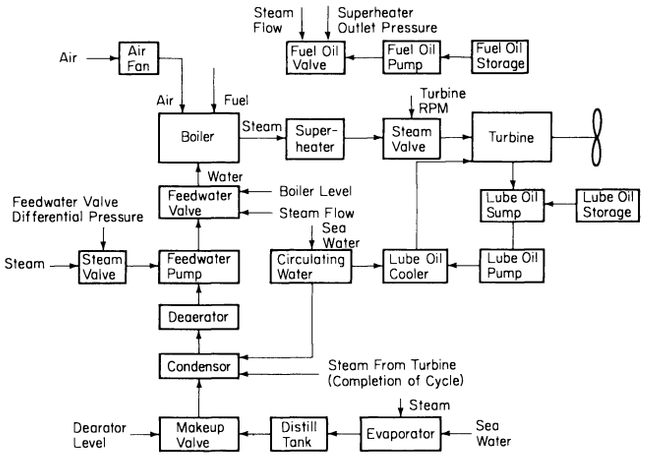

Fig. 1. Block diagram of supertanker power plant.

5) Failure of main engine r/min tachogenerator: This results in the steam valve opening to maximum and subsequently in an increase in r/min which is indicated by the auditory r/min counter. Operator should switch to manual control by using the main throttle valve to control r/min.

6) Failure of lube oil cooler tube: This results in a leakage of lube oil into the circulating water system and thus, a gradual decrease of the level in the lube oil sump tank and, eventually a "low sump tank level" alarm. Once the failure is located, the operator should stop or reduce engine speed, temporarily bypass the oil cooler, and repair it.

The subject population studied included 36 marine engineering officers who ranged in experience from four to six years. The subjects came to TNO in groups of four to six for the purpose of participating in a one-week training program in fault diagnosis, energy management, and more basic marine engineering fundamentals. For the study reported here, trainees were observed on Tuesday and Friday afternoons of the training week. During each observation period, each trainee was confronted with a single failure which he individually had to detect, diagnose, and compensate.

The following four methods were used to measure performance.

1) Verbal protocols were collected by having subjects "think aloud" and tape recording the results.
2) Computer logs of all discrete events (i.e., alarms and control actions) were printed out.
3) At first interviews, but later questionnaires, were used to assess trainees' knowledge of system operations in general and controller functions in particular.
4) Subjective ratings by instructors of a variety of dimensions of trainee performance were assessed.

These measures were used to evaluate numerous aspects of trainee performance and the TNO training program. These evaluations are described in great detail in van Eekhout's dissertation [2]. As noted earlier, this correspondence will only consider identification, classification, and interpretation of operator errors.

Classification of Errors

Before errors can be classified they must be identified. Fortunately this process was fairly straightforward. The first author, who has over ten years of experience as an engineering officer and instructor in the Dutch Navy, studied the computer logs and

TABLE I
ERROR CLASSIFICATION SYSTEM AND RESULTS

Error Categories		Failure						
General Category	Specific Category	Boiler Level Transmitter	Fuel Oil Pump	Automatic Changeover	Differential Pressure Transmitter	r/min Tachogenerator	Lube Oil Cooler	All Failures
1) Observation of system state	a) incomplete	3	1	1	6	0	0	11
	b) inappropriate	0	1	0	0	0	0	1
	c) lack	0	0	0	0	2	0	2
2) Identification of fault	a) incomplete	0	0	2	0	0	0	2
	b) inappropriate	9	4	0	5	2	2	22
	c) lack	1	0	0	1	2	3	7
3) Choice of goal	a) incomplete	0	1	0	0	0	0	1
	b) inappropriate	0	0	0	0	1	1	2
	c) lack	0	0	0	0	0	0	0
4) Choice of procedure	a) incomplete	0	0	0	0	0	0	0
	b) inappropriate	0	0	0	0	0	5	5
	c) lack	0	0	0	2	0	1	3
5) Execution of procedure	a) incomplete	0	16	3	4	0	0	23
	b) inappropriate timing	0	1	0	0	0	1	2
	c) inadvertent action	2	3	0	0	0	0	5
Contributing Factors	1) basic knowledge	1	0	0	0	0	1	2
	2) controller knowledge	8	2	2	5	4	0	21
	3) design inadequacies	3	10	0	0	2	0	15
	4) fidelity inadequacies	1	8	0	0	0	0	9

verbal protocols for each of the failures encountered by the last four groups of trainees. This included 40 sets of logs and protocols (i.e., 40 failures).

Using this data, all operator decisions and actions that could *without doubt* be identified as errors were noted. For each of these errors, a one paragraph description of *what* happened was developed. During this identification phase, no attempt was made to generalize or determine *why* a particular error occurred.

During the next phase of the analysis, each of the authors independently studied these descriptions and attempted to classify each error using the general scheme of Rasmussen which was discussed earlier in this correspondence. Unfortunately this classification scheme covers such a wide range of situations that all of the errors identified for this particular study fit into just two or three categories. Since such a high level of aggregation was undesirable, a more fine-grained version of the relevant portion of Rasmussen's scheme was developed and is shown in Table I.

The general categories in the error classification system are defined as follows.

1) Observation of system state: This occurred when an operator failed to collect appropriate and sufficient information about flows, pressures, etc., before proceeding to attempt to diagnose the failure.

2) Identification of fault: This occurred when an operator *explicitly* confirmed the wrong hypothesis or rejected the correct hypothesis.

3) Choice of goal: This occurred when an operator chose to compensate for the symptoms and ignore the cause, or chose to respond to requests from the bridge without considering the power plant's current status.

4) Choice of procedure: This occurred when an operator's choice of procedure, including informal procedures, was not consistent with his choice of goal.

5) Execution of procedure: This occurred when an operator omitted procedural steps, performed steps out of sequence, performed steps too early or too late, or committed apparently inadvertent isolated discrete actions.

Note that the categories follow a natural flow from onset of the symptoms of the failure to compensation for the consequences of the failure. Errors that are the logical results of previous errors (e.g., choosing the wrong procedure because the wrong goal was chosen) are not counted as errors.

In addition to the categories in Table I, each error was considered in terms of whether or not any of the following contributing factors appeared to be present:

1) operator lack of knowledge about the functioning of the basic system,
2) operator lack of knowledge about the functioning of the automatic control systems,
3) human factors design inadequacies of the control panel, and
4) simulator fidelity inadequacies.

The extent of the effects of the first two factors could be determined using the postexperiment questionnaires where the trainees answered multiple choice questions concerning overall system and controller functions.

Design inadequacies of the control panel were determined by a human factors analysis of the displays and controls on the panel. The primary inadequacies included

1) a plethora of auditory and visual alarms, many of which were only related to secondary effects of the failure,
2) use of closely spaced control knobs that were identical in terms of shape and color but controlled very different functions,
3) labeling of control knobs such that the labels could not be read when one's hand was on the knob, and
4) use of interlocks with confusing logic.

Estimation of simulator fidelity inadequacies was limited to the static features of the simulator and did not consider the dynamics of the process being simulated.

Using the revised classification scheme shown in Table I, as well as the potential contributing factors noted above, each author again independently attempted to classify each error. Upon completion of their independent classifications, the authors compared results and found almost complete agreement, the few exceptions being primarily related to the second author's lack of marine engineering experience.

Results

Table I summarizes the overall results of the error analysis and classification. The table entries represent the total number of errors that occurred for all subjects. There were four instances of the automatic changeover and r/min tachogenerator failures and eight instances of the other failures.

There were 86 errors for the 40 failures encountered. Thus the human error rate was approximately two per failure. However, this figure is somewhat deceptive since almost all of the errors were reversible once the operator realized his mistake. Thus only a few of the errors would have truly lead to costly consequences. Nevertheless, almost all of the errors cost the operator in terms of wasted time.

Considering the frequencies of particular types of error, the following results are notable:

1) 23 (27%) of the errors relate to incomplete execution of procedures which includes omission of procedural steps and, to a slight extent, out of sequence steps;
2) 22 (26%) of the errors relate to inappropriate identification of the failure, which includes both false acceptances and false rejections; and
3) 11 (13%) of the errors relate to incomplete observation of the state of the system prior to forming hypotheses regarding the cause of the observed symptoms.

Thus almost two-thirds of the errors fit into just three categories.

To determine the effects of the contributing factors noted in Table I, correlation coefficients were calculated for each of these factors with the occurrence of errors in each of the general categories. It was found that a lack of knowledge of basic system and controller functions was significantly correlated with the occurrence of errors in identifying the fault ($r = 0.77$, $p < 0.05$). Further, the presence of design and fidelity inadequacies was highly correlated with the frequency of errors in executing procedures ($r = 0.94$, $p < 0.01$).

Conclusion

The results presented, particularly when combined with the broad range of analyses reported in van Eekhout's thesis [2], lead to two fairly specific and important conclusions. First of all, it is not surprising that human factors design inadequacies and fidelity problems led to human errors. Nevertheless, it is important to document and quantify the extent of this problem in realistic settings such as were studied by the authors. This is particularly important because the design inadequacies of the simulator used in this study reflect the way in which the actual ship, on which the simulator is based, was designed. On the other hand, the fidelity problems appeared to be intrinsic to these types of simulator in general, rather than to this simulator in particular.

The second conclusion is more subtle. The results indicated that a lack of knowledge of the functioning of the basic system as well as automatic controllers was highly correlated with errors in identifying failures. Since most of this lack of knowledge related to controller functions (see Table I), this result points to a need for human operators to know how their automatic systems will respond in failure situations. However, most training programs appear to emphasize normal operations (e.g., controller tuning for energy efficiency) and therefore, operators seldom learn about the failure modes of automatic systems. While it may be possible to display this type of information to the operator when such failures occur, this would require a fairly intelligent computer system. Further, without appropriate training, operators might not be able to utilize this information.

In fact, if one considers the increasing use of automation for normal operations, one could argue that the emphasis of the operator's training should be increasingly shifted towards developing the human's abilities to deal with failure situations. The study reported in this correspondence indicates that an essential aspect of those abilities is being able to predict what the automatic system will do when it is not working correctly.

Acknowledgment

The authors gratefully acknowledge the use of the facilities of TNO Institute for Mechanical Constructions in Delft, The Netherlands. They are also indebted to Ir. W. B. Jaspers for his support and encouragement throughout this work.

References

[1] J. Rasmussen and W. B. Rouse, Eds., *Human Detection and Diagnosis of System Failures*. New York: Plenum, 1981.
[2] J. M. van Eekhout, "Fault detection, diagnosis, and correction by human operators in the engine room of a supertanker," Ir. Thesis, Delft University of Technology, The Netherlands, 1980.
[3] J. Rasmussen, "What can be learned from human error reports," in *Proc. of NATO Conf. on Changes in the Nature and Quality of Working Life*, Thessaloniki, Greece, 1979.
[4] J. Rasmussen, O. M. Pederson, A. Carnino, M. Griffon, G. Mancini, and P. Gagnolet, "Classification system for reporting events involving human malfunction," Riso National Lab., Roskilde, Denmark, Rep. M-2240, Mar. 1981.

The Logic of Weighted Queries

PAUL B. KANTOR

Abstract—**A mathematical formulation of the logical relationships between weighted queries is developed. It includes a unique formal element, the vapid query, which does not have any effect on the queries with which it is combined. The operations of union and intersection are defined, and shown to behave properly as the weight of the queries varies. A realization of this structure in terms of fuzzy subset functions with special rules of combination is presented. Some implications and applications are discussed.**

I. Introduction

The language of fuzzy subsets[1] has seemed, to many investigators, to offer a promising tool for the quantification and systematization of some important concepts in information retrieval.[2] The problem generally considered is a collection of objects (documents) which represent potential answers to a proposed query Q. The degree to which a single document represents an answer to Q is called its relevance to the query Q. When the relevance is assumed to take only two values (which may be thought of as zero and one) the relevance relation defines a subset of the collection of documents, that is, the subset for which the relevance to Q is one. The problem of compound queries, such as "Q and R," or "Q but not R," or "Q or R, but not both," is then mapped into the operations defined on the class of subsets of the original collection of documents. The algebra of subsets is well

Manuscript received September 12, 1980; revised August 10, 1981. This work was supported by Tantalus, Inc., Research Project No. 80.7.

The author is with Tantalus, Inc., Library Management Consultants, 3257 Ormond Rd., Cleveland, OH 44118.

[1] The concept of a fuzzy set was introduced by Zadeh [16]. Many of the points presented in that paper have since been "rediscovered" by others citing it, and the literature is enormous. A bibliography for the first decade (ending 1977) is given by Gains and Kohout [5].

[2] There is an ongoing discussion of the application of fuzzy sets to retrieval problems in the "information retrieval" literature. Tahani [14] and Radecki [9] have been particularly active. Robertson [10] has argued for a product structure which parallels probability theory. As Zadeh [16] pointed out, the product combination of fuzzy sets does not preserve the partial order structure. It also raises the problem of "When is a query a complement?" One of the referees of this correspondence calls attention to the work of Yager [15].

0018-9472/81/1200-0816$00.75