Risk and Information Processing

Rasmussen, J.

Publication date:
1985

Document Version
Publisher's PDF, also known as Version of record

RISØ-M-2518

RISK AND INFORMATION PROCESSING

Jens Rasmussen

Abstract. The reasons for the current widespread arguments between designers of advanced technological systems like, for instance, nuclear power plants and opponents from the general public concerning levels of acceptable risk may be found in incompatible definitions of risk, in differences in risk perception and criteria for acceptance, etc. Of importance may, however, also be the difficulties met in presenting the basis for risk analysis, such as the conceptual system models applied, in an explicit and credible form. Application of modern information technology for the design of control systems and human-machine interfaces together with the trends towards large centralised industrial installations have made it increasingly difficult to establish an acceptable model framework, in particular considering the role of human errors in major system failures and accidents. Different aspects of this problem are discussed in the paper, and areas are identified where research is needed in order to improve not only the safety of advanced systems, but also the basis for their acceptance by the general public.

.../...

INIS Descriptors. DECISION MAKING; FUNCTIONAL ANALYSIS; HUMAN FACTORS; INDUSTRIAL PLANTS; INFORMATION NEEDS; MAN-MACHINE SYSTEMS; NUCLEAR POWER PLANTS; PLANNING; RISK ANALYSIS; SYSTEM FAILURE ANALYSIS.

August 1985
Risø National Laboratory, DK-4000 Roskilde, Denmark

A satisfactory definition of "human error" is becoming increasingly difficult as the human role in systems is changing from well trained routines towards decision making during system malfunctions. Recent research on the cognitive control of human behaviour indicates that errors are intimately related to features of learning and adaptation, and neither can nor should be avoided. There is, therefore, a need for design of more error-tolerant systems. Such systems depend on immediate recovery from errors which, in turn, depends not only on access to factual information about the actual state of affairs, but also on access to information about goals and intentions of planners and cooperators. This information is needed as reference for judgements, but is difficult to formalise and is not at present included in interface and communication systems to any large degree. As the information systems are becoming more "intelligent" and systems for cooperative decision making are being designed, the influence of the users' understanding and acceptance of advice from a computer will be critical for overall risk from large-scale system operation.

Invited paper presented at Seminar on Risk: Decisions, Concepts, and Measures. Norvegian Risk Research Committee, June 17-19, 1985. Oslo, Norway.

ISBN 87-550-1138-1
ISSN 0418-6435

Grafisk Service Center 1985

TABLE OF CONTENTS

INTRODUCTION

Two trends in the industrial and technological development have had major impacts on the problems in coping with the risk involved in industrial operations. First of all, there has been a general trend towards large and centralised operations, not only in production plants but also in administrative systems, commercial companies and outlet chains, with the consequence that faults and errors can lead to drastic damage and economic loss. Examples from recent years are legio. This situation has immediately two consequences in the present context. On one hand, during system design it is now becoming necessary to consider the consequences of events and conditions of very low probability. On the other, due to the short time span between conceptualisation of new products or processes and full-scale operation, this cannot be done from direct empirical evidence or operation of prototype systems. During periods of rapid development and with changes of basic technologies, piecemeal adjustment of prior designs is no longer adequate. Instead, new analytical methods have to be found, and a "top-down" design approach based on proper predictive models is necessary. Such a design approach has to include a consideration of the ultimate risk related to operation by means of systematic analytical risk assessment.

This industrial development has led to a widespread public concern with the safety of such installations, and the designers have made serious attempts to explain and document the safety targets underlying the design and the probabilistic considerations by which the design is validated. Such attempts have had limited success, with the consequence that the difference between objective risk concepts of system designers and the subjective risk perception of the general public has been widely studied and discussed. Based on the assumption that quantitative risk figures are not understood by the general public, designers have made great, but largely unsuccessful attempts to compare their risk figures with other categories of natural and man-made risks, and many attempts have been made to explain the lack of accept in terms of difference in risk acceptance depending on the degree of voluntary exposure, in acceptance of individual or collective risk, etc. This approach in turn being based on the assumption that a kind of more or less conscious risk evaluation is underlying the different personal choices.

This argumentation may be misleading in two respects. First of all, the lack of accept of the risk figures resulting from quantitative risk assessment may not only be related to acceptance of the risk level per se, but also to a lack of confidence in the underlying assumptions of the analysis. Secondly, the concept of risk cannot be separated from other aspects of personal value judgements underlying intuitive human choice.

In the following sections, it is argued that the present rapid development of information technology may increase the difficulty in formulating a credible basis for risk analysis and point to areas where basic research is needed in order to improve the model framework behind risk analysis.

TECHNOLOGICAL TRENDS

Analysis of major accidents has invariably shown that human activities have been involved in the causation and further development of the course of events. Reviews show that human errors have been significant in typically 70 to 80% of accidents reported. Coping with the human role is clearly important, but the new technology is presenting special problems since analytical assessment requires predictive models of human performance, models which are not presently well developed.

Another important industrial trend is the introduction of modern information technology in the interface between humans and their actual work content. Information technology will now allow, both technically and economically, the design of very complex control systems. This means that all frequent and, therefore, well formulated tasks are likely to be automated. Left to the human supervisor are the more creative tasks related to problem solving. In consequence, system design cannot be based on the traditional analysis and description of tasks and functions. Instead, design should be in terms of an envelope within which an operator can adopt effective strategies during situations which have only been foreseen by the designer in kind, not in particular. Modern computers and information displays are effective means for system designers in their attempts to match the selection and presentation of information to the various human tasks and roles. This optimisation cannot be based on traditional task analysis in terms of action on the system, but an

analysis in terms of cognitive or mental activity is necessary. Misconception in this aspect of design may leave the human in a situation which is worse than it would be by less advanced technology. In consequence, methods for analytical risk assessment including the human activities in system operation and maintenance become for design of large-scale industrial installations.

OBJECTIVE AND SUBJECTIVE RISK

The basic risk concept involved in the objective evaluations applied by designers and in the subjective judgements of the general public is in both cases related to an aggregated measure of probability and magnitude of negative effects. The actual decision processes are ,however, basically different. The evaluation performed by designers is based on a conscious, analytical comparison of quantitative measures of separate aspects of costs and benefits, such as productive output and losses caused by disturbances.

Most human choice is not, however, based on such rational analysis which is mainly typical of formal professional activities, but on holistic, intuitive and mostly subconscious judgements. Risk will be only one feature of the value perception underlying intuitive choices, and will not be considered separately, except maybe for after-the-fact explanation. From research on social judgement it is wellknown that judges use fewer, and frequently different cues than they can rationalise when asked. The context in which risk may influence judgement will vary widely with the particular setting. In general public acceptance of new technologies, such as, for instance, nuclear energy or genetic engineering, value judgements also involve features of general political nature and emotional reactions to unknown technologies. In questions whether to choose living close to such installations, features of immediate quality of life aspects interfere. When entering activities like skiing, mountain climbing, or high speed driving, risk perception probably has much less weight in judgements than the immediate feedback from perception of the limits of control which is necessary for improvement of skill, together with the joy experienced from such skill improvement. Considering choice of the individual acts of an activity, risk considerations may be involved during

conscious preplanning, but when once absorbed in the performance of a skilled task, our analysis of accident reports and verbal protocols indicates activity to be controlled by immediate functional criteria like "choosing the way of least effort" and control by higher level features like risk potential appears less likely.

The role of risk features in a judgement, therefore, cannot be separated from the context by objective analysis, and rational arguments based on comparison of the perception of risk across activities and situations cannot be expected to change people's actual judgements.

An important reason for the lack of acceptance of risk in spite of the objectively acceptable quantitative figures may be the lack of confidence in the assumption underlying risk analysis, a criticism which is difficult to formulate by the general public, and which may, therefore, be expressed in terms of risk level.

One major problem in analytical risk assessment is to obtain a clear and explicit formulation of the coverage of an analysis. The final result of a risk analysis is a theoretical construct which relates empirical data describing functional and failure properties of equipment and processes to a quantitative or qualitative statement of the overall risk to be expected from the operation of the system. The analysis depends on a decision regarding the boundaries of the system to be considered; on a model describing the structure of the system and its functional properties in normal and in all relevant abnormal modes of operation, including the activities of the people present in the system; together with a number of simplifying assumptions which are necessary to facilitate a systematic analysis. These assumptions, the model, and the characteristics of the sources of the empirical data used, are just as important results of the assessment study as is the resulting risk figure, since they are the necessary precondition for the operation of the system in correspondence with this risk target.

Unfortunately, neither this basis of analysis nor the strategies used to identify the mechanisms and courses of events to include in the analysis, are generally explicitly formulated in present analytical techniques. This makes it difficult to use the analyses for control of the actual conditions during operation. The conclusion is that a major problem in presentation of credible risk analysis is the formulation of the underlying models and assumptions in a way which makes possible an independent

verification of the correspondence with the actual installation. The aim of the present paper is to point to several aspects in the technological development which aggrevate this problem and, therefore, require more active research efforts, in particular with respect to models including human activities and the effects of modern information technology.

NATURE OF HUMAN ERROR

The trends discussed in the introduction invite a closer look on the nature of human errors. In the industrial or technical context, the definition of a human error has typically been made in analogy to component faults. For analytical risk assessment, a technical installation has been considered an aggregation of standard components for which failure characteristics and frequencies could be determined empirically from application in other systems. The overall risk involved in system operation can then be calculated or simulated by means of a model of the causal structure of the system. In analogy, human performance was considered an aggregation of standard acts or routines for which error characteristics and frequencies could be collected from similar activities in other task settings.

This approach has close links to Taylorism in industrial engineering and behaviourism in psychology, and has been fruitful in analysis of systems where human activity has been manual assembly tasks, repair, and calibration. Such tasks can often be decomposed into more or less separate, manual routines, and analysis can be based on the overt activity which to a large degree is controlled and sequenced by the system. Another important feature is that many tasks have been repetitive, and that performers have reached a stable level in a skill in which errors can be considered stochastic variations going beyond the limits acceptable for proper system performance.

The application of modern information technology is rapidly changing the basis for these assumptions. Automation has removed many repetitive tasks and given humans the role of supervisors and troubleshooters. This means that their performance is more related to decision making and problem solving and, consequently, cannot be adequately decomposed into standard routines in terms of overt, observable elements. Analysis has necessarily

to be performed in terms of cognitive information processing related to diagnosis, goal evaluation, prioritising, and planning. Such mental functions are much less constrained by the external task conditions. They can be solved successfully by several different strategies and the individual choice will depend on very subjective criteria. Another important point is that performance in a task can no longer be assumed to be at a stable level of training. Learning and adaptation during performance will be significant features of many situations which are relevant for analysis of the ultimate risk. It follows that analysis of the human role in this risk can no longer be based on a model of the external characteristics of the task. Design has to be related to a model of human performance in psychological terms referring to cognitive capabilities and limitations.

Furthermore, when performance can no longer be judged with reference to a stable normal performance, the definition of "human error" becomes dubious. Considering a highly skilled performance of a task there will generally be no difficulty in identification of errors and no dispute between the performer considering his actual goals and intentions and a posterior analysis. However, considering performance during complex abnormal situations which are part of an accidental scenario there is no clear reference for identification of "errors". They are found during the search for causes of the accidental event, but the identification in terms of component fault, operator error, manufacturing error, or design error depends entirely upon the stop-rule applied for termination of the search. This stop-rule will be purely pragmatic and be something like: An event will be accepted as a cause and the search terminated if the causal path can be followed no longer, **or** a familiar, abnormal event is found which is therefore accepted as explanation, **and** a cure is known. Paradoxically, human errors seem to be allocated under two typical circumstances. On one hand, human errors are found when human variability brings an otherwise stable task outside acceptable limits. On the other, human errors are found when human variability or adaptability proved insufficient to cope with variations in task content; if, on hindsight, a "reasonable" human ought to be able to cope with disturbances.

It appears to be a more fruitful approach not to look for errors as causes of accidents, but to consider the related events to be occasions of human-task mismatches and to look for factors which are sensitive to improvement, whether or not they are considered causes, i.e. irrespectively of their location on the causal

path. Accidents can be avoided by breaking the path, as well as by removing causes, as everybody will know (Leplat and Rasmussen, 1984).

The nature of the tasks in modern systems, being related to problem solving and decision making in which adaptation to unfamiliar situations is crucial, makes it very doubtful whether a category of behaviour called errors can be meaningfully maintained. The term "error" in a way implies that something could be done to the humans in order to improve the state of affairs. Recent work on the problem indicates that effective means can more readily be found when considering design of "error-tolerant" systems - by means of modern information technology.

Basically, system designers have to accept human variability as an integral element in human learning and adaptation (Rasmussen, 1984). Fine-tuning of manual skills depends upon a continuous updating of the sensory-motor schemata to the time-space features of the task environment. If the optimisation criteria are speed and smoothness, adaptation can only be constrained by the once-in-a-while experience gained when crossing the precision tolerance limits, i.e. by the experience of errors or near-errors. These, then, have a function in maintaining a skill, and they neither can nor should be removed. Also at the more consciously controlled rule-following level, development of know-how and rules-of-thumb is depending upon a basic variability and opportunity for experiments to find shortcuts and identify convenient and reliable signs which make it possible to recognize recurrent conditions without analytical diagnosis; in short, to develop quasi-rational heuristics. Involved in genuine problem solving, test of hypothesis becomes an important need. It is typically expected that operators check their diagnostic hypothesis conceptually - by thought experiments - before operations on the plant. This, however, appears to be an unrealistic assumption, since it may be tempting to test a hypothesis on the system itself in order to avoid the strain from reasoning in a complex causal net. For this task, a designer is supplied with effective tools like experimental set-ups, simulation programs and computational aids, whereas the operator has only his head and the plant itself. And - "The best simulation of a cat - is a cat." In this way, acts which on afterthought are judged to be mistakes, may very well be reasonable acts intended to gain information about the actual state of affairs.

In other words, considering the human role in modern systems, human errors should rather be considered to be "unsuccessful

experiments in an unfriendly environment", and design efforts should be spent on friendly, i.e. error-tolerant, systems.

The view that "errors" are integral parts of learning mechanisms has long roots. Already Ernest Mach (1905) notes: "Knowledge and error flows from the same mental sources, only success can tell the one from the other", and Selz (1922) found that errors in problem solving were not stochastic events, but had to be seen as results of solution trials with regard to the task, which is somewhat misconceived. Hadamard, the mathematician, states (1945): "-- in our domain, we do not have to ponder with errors. Good mathematicians, when they make them, which is not infrequent, soon perceive and correct them. As for me (and mine is the case of many mathematicians), I make many more of them than my students do; only I always correct them so that no trace of them remains in the final result. The reason for that is that whenever an error has been made, insight - that same scientific sensibility we have spoken of - warns me that my calculations do not look as they ought to".

This means that human errors cannot be studied in isolation, only as a part of an analysis of the psychological mechanisms controlling cognitive activities in general. Only quite recently has research in cognitive psychology again taken up the interest in such studies (Reason, 1982; Norman, 1980). The findings match very well those found from analysis of industrial accidents (Rasmussen, 1980), and indicate that the great variety of errors can to a large degree be explained as the effect of a very limited number of psychological mechanisms when folded onto the variety of the work environment - as Simon (1969) argues: --"man is quite simple, complexity of his behaviour reflects largely the complexity of the environment."

DESIGN OF ERROR-TOLERANT SYSTEMS

It follows that system designers have to accept that humans make errors all the time, and that this is just the other side of the generally successful adaptation of the user's behaviour to the peculiarities of the system. Or, as Reason has said it : "Systematic error and correct performance are two sides of the same coin" (Reason, 1985). The task of the designer will be to aim

for error-tolerant systems, in which errors are observable and can be reversed before unacceptable consequences develop.

This brings the use of advanced information technology in the work interface into focus. It is now possible to match the interface to the requirements of individual users and their immediate tasks. From a risk point of view this may lead to problems, considering the importance of certain categories of rare events for the safety of many kinds of systems. Optimising an interface to the requirements of the more frequent members of the task repertoire for which performance can be evaluated empirically may create difficulties in more unfamiliar task situations. Furthermore, optimising for support of task execution may violate requirements from error recovery.

Task execution is based on procedural information of the form: If (situation, cue), then do (action, task). For familiar tasks, this information is immediately available in terms of skilled users' know-how, and computer support of the less skilled can be developed in computerised procedure retrieval systems or, in more recent terms: expert systems. For monitoring the effect of the activity, and for recovery from disturbances, quite a different kind of information is needed. Error detection is not simply a question of monitoring the outcome in comparison with the goal. In many cases this will lead to detection far too late - you cannot save the cake when tasting the final result. Monitoring depends on the equivalent of Hadamard's "scientific sensibility" which is something like understanding of the functioning of the system behind the task and knowledge of the intended dynamic behaviour. It is important to understand and to monitor the process, not only the product. Of major concern from a risk point of view, when attempts are made to transfer the heuristic rules of "know-how" of human experts to "expert systems", should be the problems in transferring also the "sensibility" to the limits of expertise. The applicability of "expert systems" in centralised systems very much depends on the ability to appeal to analytical performance when the preconditions for heuristical rules break down (Barnett, 1982).

In other words, monitoring of routine activities probably depends on the same kind of information as diagnosis and intervention during infrequent tasks: ability to predict the behaviour of the system and to compare with the intended performance. This additional need should be carefully considered in the design of interface systems.

CAUSES AND REASONS

Decision making is in general a kind of resource allocation in a problem space which can be organised in a means-end hierarchy reflecting the fact that the system involved can be described at several different levels of abstraction in the mapping of the purpose/function/equipment relationships. A decision task involves the identification of discrepancies between the actual state of affairs and the target states, which may be done at any of the levels in the means-end hierarchy. In this hierarchy, the effects of changes in the physical world propagate bottom-up, and the reasons for proper functions are derived top-down. In the design of information systems, emphasis is typically placed on representation of factual information from measurements and statistics, i.e. bottom-up data. This is due to an assumption that decision making for supervisory plant control as well as executive management depends on rational analysis of the system involved and is performed in accordance with the formal, theoretical decision models. The information about purposes, reasons and policies is only implicitly formulated; it is assumed to be available from general training and instruction. This may be the case for undisturbed routine tasks, but for infrequent tasks and for error detection it is not necessarily true. When introducing information systems as an interface to the task content, the disturbance of informal top-down paths for communication of reasons should be carefully considered.

A decision making task which has been in the focus of discussion during the recent decade has been that of industrial operators during system failures. In industrial process plant control rooms, a large amount of measured plant status data are presented to the operators and great effort is spent on development of proper presentation of this information and support of the operators in diagnosis, i.e. bottom-up identification of the actual physical state of the plant. In addition, support of the operators' memory of the functional structure of plant is given in terms of mimic diagrams, etc. Operators are generally supposed to assess the operation of the system from understanding of the functional structure and knowledge of physical variables. This is frequently not the case. Many systems, for instance control systems, are too complex and operators will rather try to judge correctness of function with reference to their perception of the designer's intentions, i.e. from information derived top-down in the hierarchy.

At present, data bases for industrial control rooms include only little information about the complex relationship between overall purposes and goals and the intentions behind the design at the lower levels of functions and equipment. This is so, partly because such information is difficult to formalise, but also because it is only implicitly present in the form of company policies, design practices, and system designers' subjective preferences which do not find their way into drawings and technical manuals. Instead, reliance has been on ad hoc advice facilities, e.g. supervisors on call, communication with designers, etc. In the nuclear industry, great effort has been spent in formalising such systems in terms of "resident technical advisors", "technical support centers", data links to design teams and authorities. There is, however, a movement towards exploitation of advanced information systems, "expert systems", for such advice giving, and direct transmission of plant status data to outside advisors by data links is considered. At the same time, there is a tendency towards an integration of the process computer systems and the computer systems used for production and maintenance planning. This integration of plant control and executive decision making may have implications for risk management, if based on a conception of decisions transferred from normative theories.

NORMATIVE MODELS AND EMPIRICAL EVIDENCE

In general, there is a discrepancy between the normative theories for management decision making which are typically derived from economics, and the empirical evidence. It is generally assumed that the decisions of high level executives are based on careful analysis of statistics and factual reports, the kind of information which is normally considered for computerised management data bases. Several studies indicate that this is an unreliable assumption. Dreyfus and Dreyfus (1980) find that the normative, theoretic models of decision making are only representative of the behaviour of novices, and Minzberg (1973) concludes from a study that top level executives prefer face-to-face interaction and even hearsay and gossip to analysis of factual reports. A reasonable explanation may be that management executives are not faced with a causal system, the response of which can be predicted bottom-up by factual analysis. They are actors in a social game and predictions have to consider

intentions and motives of other people rather than objective facts. Predictions have to be derived top-down from a reliable perception of other people's value structures, for which face expressions and gossip may be more reliable sources than statistical reports.

Communication of values and intention is not only required for strategical planning. It is a precondition for the error correction features which seem to be inherent in social organisations. Cyert and March (1963) call it "bias discount". When people are making frequent errors, one could fear that errors would propagate willingly in a social system and add up until a major mistake is at hand. However, this appears not to be the case. The individual agents are correcting faults in messages and data and will complete ambiguous orders and instructions from their implicit knowledge about policies and other people's intentions and goals.

Therefore, failure-tolerant management systems basically depend on the continuous and efficient communication of corporate and individual values and intentions. One of the major risk problems related to the intrcduction of information technology in centralised systems may therefore be the temptation of rational, scientifically minded experts to design large systems in which centralised data banks with factual information are the basis for communication between decision makers, and, unwillingly, disturb the communication of values and intentions which is necessary for error recovery.

There may, however, be another dimension to the communication of value structures. Such communication is crucial for error recovery but may lead to a very tightly coupled system with short time constant and, consequently, stability problems. Losses and time delays are fundamental tools for maintaining stability in technical systems. Are similar measures now necessary in social systems? The consequences of effective communication of attitudes and values can be seen at a grand scale, as a consequence of the effective communication of values by the mass media. It took the French revolution half a decade to initiate a change in Denmark, where a new constitution was the result of a meeting. The student revolt in Berkely was, however, followed next morning in Copenhagen. Small-scale experiments and adjustment of approaches at a reversible level may be difficult in tightly coupled, fast systems. Is it now necessary to consider stability theory of social systems on a control theoretic basis? If so,

approaches like Forrester's (1971) modelling should be supplemented with models of propagation of values.

ADVICE ACCEPTANCE

Closely connected to this reliability problem is the problem of advice acceptability. When several people cooperate in decision making, they will be exchanging messages communicating factual information, results of analysis, and plans for action. The form of the messages may vary, depending upon the role of the participants and upon their authority. For data, the form ranges from statements of facts to hypotheses of varying likelihood, and plans for actions may be stated as proposals, advices, instructions, or orders. It is crucial for the reliability of cooperative decision making that messages are received in the mode they were intended by the sender, e.g. the designer of a computer-based decision support system. For orders, this is no great problem, but the question of criteria for proper understanding and acceptance of advice and recommendations is crucial. What kind of information is needed for making an advisee understand an advice properly? The problem has been discussed in some detail for "expert systems" for support of medical diagnosis like MYCIN (see for instance Shortliffe, 1983), but the solution proposed, which is a replay of the inference rules used by the advisor, does not appear to be convincing. Understanding of a piece of advice depends not only on a step-by-step tracing of the way in which the result was found, but also on reasons why that path was chosen. It is important to consider that human decision makers are quasi-rational; underlying analytical reasoning there is a background of intuitive judgement and expectations. The composition of intuition and analysis depends entirely on the familiarity of the problem context to the decision maker and, consequently, so does the kind of information required to make advice understood.

The interaction between user, computer, and designer changes in a very important way when the routine tasks are automated and only the ad hoc, on-line decision making is left as an interactive task. For frequent tasks, a "task allocation" can be made. From empirical evidence the human will have definite expectations about the automated functions and intuitively be able to "understand" them. He will not need conscious, analyti-

cal evaluation of the automatic functions, he will be allowed to forget details, reasons, and necessary preconditions. This is not the case for interactive decision making where the computer is supposed to take over the data collection, preprocessing, and transformation during various phases of the decision process. The human user is then supposed to accept the result from the computer and to take over information processing during particular - and frequently badly structured - parts of the decision process. This presupposes, however, that the human accepts the immediate results of the computer, and this in a situation when the human may have no intuition and no well structured expectations to the message. The user will need to evaluate the reliability of the message in some way and will need an explanation which is not just a replay of the algorithm, but information matching the user's intuitive expectations. If this is not possible, a mode of competitive rather than cooperative interaction may develop. In such systems, design is not a question of task allocation, rather a question of allocation of authority; the task is performed more or less in parallel by the user, the computer, and the programmer. What is shifting is the role as performer, monitor, and advisor, and, with that, the mode of processing applied.

ETHICAL QUESTIONS OF DESIGN

The difficulty in highly automated systems of establishing a clear reference in terms of "normal behaviour" raises some problems for designers of large-scale systems in terms of compatibility between their expectations to user behaviour, the actual behaviour, and a posterior judgement of behaviour in case of accidents. The allocation of guilt after the fact depends on a concept of a "reasonal person" which may be very different for behaviour in very familiar situations and in case of problem solving during disturbed situations. This "reasonable person" at times seems to be rather similar to the normative decision theorist's "rational agent", and an interaction among professionals from systems design, psychology, sociology, and legal matters might be useful to probe the need for changes in the perception of human errors.

CONCLUSION

The conclusion of this discussion will be that the present rapid technological development, in particular within information technologies, makes it increasingly important to realise that conditions for systems design have changed. Up to now, systems design and planning of human work conditions have been considered two independent activities on each side of a man-machine interface which is taken care of by human factors specialists. It is symptomatic that the International Federations of Automatic Control (IFAC) and Information Processing (IFIP) have had two committees. One on "Social Effects of Automation" taking care of systems considered as the work environment of humans, and another on "Modelling Man-Machine Interaction" considering humans as functional systems components. A consequence of the tight coupling of the activities of humans and computers at the intellectual level will be that this separation is no longer possible. Human values and attitudes will not only be a question of quality of working life, but directly influence functional effectiveness and reliability. It also means that proper design is no longer a question of having practitioners in Human Factors to use the available results from academic research; no acceptable design model of higher level intellectual processes and of affective states is as yet available, and a change in academic research is needed towards analysis of complex man-machine systems in cognitive psychology, linguistics, semiotics, etc. Furthermore, it will be mandatory that researchers within these fields have a solid basis in technological knowledge and understanding. As the sociologist Peter Winch (1958) noticed:"A sociologist of religion must himself have some religious feeling if he is to make sense of the religious movement he is studying and understand the considerations which govern the lives of its participants". It will, in the same way, be impossible to study human interaction with technical systems without fundamental knowledge of the technology. This is particularly so for the design of high risk man-machine systems.

There are, fortunately, several signs of changes in the proper direction. University faculties are discussing the plans for technical-humanistic lines of education, and programs for psychological experiments in complex decision making situations are taking over interest from the classical experimental psychology paradigm. Also, committees like those of IFAC/IFIP are mutually trying to reach an integrated view of the criteria for systems design.

REFERENCES

Barnett, J. A. (1982): Some Issues of Control in Expert Systems. Proceedings of the International Conference on Cybernetics and Society. Seattle, October 28-30, 1982, pp. 1-5.

Cyert R. M. and March, J. G. (1963): A Behavioural Theory of the Firm. Prentice-Hall.

Dreyfus, S. E. and Dreyfus, H. E. (1980): A Five Stage Model of the Mental Activities Involved in Direct Skill Acquisition. ORC-80-2 Operations Research Center, University of California Berkeley.

Forrester, J. W. (1971): World Dynamics. Wright-Allen Press.

Hadamard, J. (1945): The Psychology of Invention in the Mathematical Field. Princeton Univ. Press.

Leplat, J. and Rasmussen, J. (1984): Analysis of Human Errors in Industrial Incidents and Accidents for Improvements of Work Safety. Acc. Anal. & Prev. Vol. 16, No. 2, pp. 77-88.

Minzberg, H. (1973): The Nature of Managerial Work. Harper and Row.

Mach, E. (1905): Knowledge and Error. English Edition: Reidel, 1976.

Normann, D. A. (1980): Errors in Human Performance. Report No. 8004, Center for Human Information Processing, Univ. of California, San Diego.

Rasmussen, J. (1980): What Can Be Learned from Human Error Reports? In: Duncan, Gruneberg, and Wallis (Eds.): Changes in Working Life. John Wiley.

Rasmussen, J. (1984): Human Error Data. Facts or Fiction? 4th Nordic Accident Seminar. Rovaniemi, Finland. Risø-M-2499.

Reason, J. (1982): Absent-Minded? Prentice-Hall.

Reason, J. (1985): General Error Modelling System (GEMS). In: Rasmussen, Duncan and Leplat (Eds.): New Technology and Human Error. Proceedings of 1st Workshop on New Technology and Work. Bad Homburg. John Wiley. To be Published.

Selz, O. (1922): Zur Psychologie des Productiven Denkens und des Irrtums. Friederich Cohen, Bonn.

Shortliffe, E. H. (1983): Medical Consultation Systems: Design for Doctors. In Sime and Coombs (Eds.): Designing for Human-Computer Communication. Academic Press.

Simon, H. A. (1969): The Sciences of the Artificial. MIT Press.

Winch, P. (1958): The Idea of a Social Science. Routledge and Kegan Paul.

Risø National Laboratory **Risø - M -** 2518

Risø - M - 2518

Title and author(s)

Risk and Information Processing

Jens Rasmussen

Date August 1985

Department or group

Electronics

Group's own registration number(s)

R-4-85

20 pages + tables + illustrations

Abstract

The reasons for the current widespread arguments between designers of advanced technological systems like, for instance, nuclear power plants and opponents from the general public concerning levels of acceptable risk may be found in incompatible definitions of risk, in differences in risk perception and criteria for acceptance, etc. Of importance may, however, also be the difficulties met in presenting the basis for risk analysis, such as the conceptual system models applied, in an explicit and credible form. Application of modern information technology for the design of control systems and human-machine interfaces together with the trends towards large centralised industrial installations have made it increasingly difficult to establish an acceptable model framework, in particular considering the role of human errors in major system failures and accidents. Different aspects of this problem are discussed in the paper, and areas are identified where research is needed in order to improve not only the safety of advanced systems, but also the basis for their acceptance by the general public.

A satisfactory definition of "human error" is becoming increasingly difficult as the human role in systems is changing from well trained routines towards decision making during system malfunctions. Recent research on the cognitive control of human behaviour indicates that errors are intimately related to features of learning and adaptation, and neither can nor should be avoided. There is, therefore, a need for design of more error-tolerant systems. Such systems depend on immediate recovery from errors which, in turn, depends not only on access to factual information about the actual state of affairs, but also on access to information about goals and intentions of planners and cooperators. This information is needed as reference for judgements, but is difficult to formalise and is not at present included in interface and communication systems to any large degree. As the information systems are becoming more "intelligent" and systems for cooperative decision making are being designed, the influence of the users' understanding and acceptance of advice from a computer will be critical for overall risk from large-scale system operation.

Available on request from Risø Library, Risø National Laboratory (Risø Bibliotek), Forsøgsanlæg Risø),

Copies to