Conceptual Models in Man-Machine Design Verification

Rasmussen, Jens

Publication date:
1985

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

RISØ-M-2520

CONCEPTUAL MODELS IN MAN-MACHINE DESIGN VERIFICATION

Jens Rasmussen

Abstract. The need for systematic methods for evaluation of design concepts for new man-machine systems has been rapidly increasing in consequence of the introduction of modern information technology. Direct empirical methods are difficult to apply when functions during rare conditions and support of operator decisions during emergencies are to be evaluated. In this paper, the problems of analytical evaluations based on conceptual models of the man-machine interaction are discussed, and the relations to system design and analytical risk assessment are considered. Finally, a conceptual framework for analytical evaluation is proposed, including several domains of description: 1. The problem space, in the form of a means-end hierarchy; 2. The structure of the decision process; 3. The mental strategies and heuristics used by operators; 4. The levels of cognitive control and the mechanisms related to human errors. Finally, the need for models representing operators' subjective criteria for choosing among available mental strategies and for accepting advice from intelligent interfaces is discussed.

INIS Descriptors. DECISION MAKING; FUNCTIONAL ANALYSIS; HUMAN FACTORS; INDUSTRIAL PLANTS; MAN-MACHINE SYSTEMS; MONITORING; PLANNING; RISK ANALYSIS; SPECIFICATIONS.

August 1985

Risø National Laboratory, DK-4000 Roskilde, Denmark

Invited paper presented at 1985 IEEE Third Conference on Human Factors and Power Plants, June 23-27, Monterey, LA.

ISBN 87-550-1141-1
ISSN 0418-6435

Grafisk Service Center 1985

TABLE OF CONTENTS

INTRODUCTION

The recent rapid change in the technology applied for the interface between industrial process plants and the operating personnel, combined with the efforts to support operators in their supervisory control task during infrequent, but risky plant conditions, have led to a widely recognised need for methods to evaluate the performance of new system concepts. Managers responsible for the economy and safety of large-scale industrial installations are quite naturally asking for "hard data" to prove the benefit to gain from new systems, and authorities need confirmation that the new systems will be able to meet the requirements of their regulations.

The search for methods for empirical evaluation of new decision support systems by means of laboratory experiments or simulation of critical functions of the total system has not had definite results, and not infrequently is one left with the impression that evaluations are afterrationalisations of the judges' intuitive hunches rather than objective data. The same system may very well be judged good or inferior in consequence of differences in the - often implicit - preconditions of the tests. This statement does not imply that evaluators are dishonest or have hidden motives, it is simply a consequence of the well-known confirmation bias and of the fact that an evaluator will only have reasons for continuing the analysis of preconditions of his data as long as he has some doubts (research did not lead to whole number atomic weights until theory told that broken numbers were probably due to measuring errors (Kuhn, 1962)).

The evaluation of a design is often considered a separate process which should serve as a link between the conceptual framework of the designer and the needs of the end-user; a kind of independent guarantee that concept and needs will match. This can be the case for systems for which the performance requirements are primarily related to the efficiency of normal, everyday performance. For interface systems in large-scale industrial process systems this is typically not the case. Important requirements are related to the performance during very infrequent and unfamiliar task conditions. The conclusion of the present discussion will be that during a transition phase of the interface technology applied in high risk systems, evaluation cannot be a separate process. An intimate interaction is required between system design, plant staffing, training, and empirical tests, concurrent with more basic psychological experiments.

DESIGN AND EVALUATION

In theoretical discussions, the conception of design is typically a more or less orderly process leading from the statement of goals, through several levels of functional formulations, to the final choice of the material implementation. In general, iterations between phases are included, but by and large the top-down, "vertical" design process is adopted as the formal model. This view does not correspond to realities during periods of rather stable technology. Design is then largely a "horizontal" process. Previous designs are updated to incorporate new components or to respond to new user requirements within the established overall framework, and the conceptual basis of - the reasons for - the established practice may no longer be explicitly formulated. However, indepc 'ent evaluation is no major requirement, since designers and users share this established practice as a basis for discussion of the merits and difficulties to be expected from proposed changes. Evaluation by "expert judgement" may bring you a long way.

During periods of major technological changes, this established practice should ideally be replaced by a more explicit formulation of the goal-function-implementation relationship in a systematic top-down design process. This is, however, difficult, since the user and the designer will have no established common framework for the formulation of goals and functions. With major changes in technology and tools, the needs and work organisation of users change in a way which may be unpredictable even for themselves, and the specifications necessary for a systematic design may be an unrealistic demandAt the same time, design is not a typical part of engineering curriculae, which are mostly strong in the formal analysis of systems which are themselves the results of "the art of design". The odds are that the coming successful designs will be based on inventions, intuitive conceptions, etc., and the development will be controlled by industrial competition and survival of the "fittest design".

This means, however, that a process of "evaluation" which will then serve to make the conceptual basis of a design explicit and to test the match to the need of potential users is an indispensable requirement.

EVALUATION

It is generally accepted that design evaluation is a complex process which can be approached from many different points of view. Hollnagel (1981) discusses several aspect of evaluation from the point of view of hypothesis testing in experimental psychology, and in particular emphasises the need to consider two aspects in design evaluation: The verification of the correspondence of the final system with the conceptual system as intended by the designer, and the validation of the correspondence of the design concept with the conditions of the real world of application.

To these considerations, Rouse and his collaborators (Rouse et al., 1984a&b) add the discussion of empirical versus analytical evaluation. The discussion of Rouse et al. presents the most systematic analysis of the methodological problems of design evaluation of decision support systems in process plants. The approach includes a classification of the decision support systems, based on a generic model of the elements of a decision task. This step serves as an afterrationalisation of the design process in order to cope with the frequent lack of reference for verification due to the implicit nature of most designs, mentioned above. Their approach also includes a test of the internal functional consistency of a design in the form of a top-down analytical evaluation, during which a test is made to ensure that three main aspects of the design are appropriately considered during design: The effectiveness, i.e. whether messages presented to users contain the information required for the functions identified by the classification analysis. The understandability, i.e. whether a user will be able to understand the message, given the actual task context and training. And finally, the compatibility should be considered, i.e. the physical presentation of messages and the responses expected should match the input-output capabilities of the users. Although these analyses should in fact be an integrated part of the design, the general experience is that they are needed as a part of the evaluation process to prepare the references for judgement during the empirical evaluation and to screen for inconsistencies in advance of the experimental efforts. As it is argued by Rouse et al. the rational approach to an empirical test is bottom-up, starting with test of compatibility and understandibility, before the complex testing of effectiveness is prepared. It serves no purpose to plan the complex testing of the effectiveness of the total system concept, if the individual messages cannot be immediately understood. The empirical evalu-

ation will generally involve a series of experimental sessions of growing complexity starting with simple legibility tests or checklist walk-throughs, through part task simulations, to large-scale high fidelity simulator scenarios. This systematic approach will greatly improve the benefit to be expected from attempts to evaluate new designs, in particular for systems in which the performance during the normal work conditions is important. However, difficulties are still to be expected when the evaluation of performance during infrequent and unfamiliar situations is vital. Several large-scale simulator experiments have been performed in order to evaluate new decision support and disturbance analysis systems (Woods et al., 1981). Rather than giving clear-cut evidence on the benefit to be gained by the new systems, these experiments have given very important data for the understanding of human behaviour during unfamiliar task conditions which is necessary for analytical evaluations. One major problem, which is not solved by the presently available approaches to empirical evaluation, is the selection of subjects for large-scale simulation of infrequent events in new systems. Well-trained users from previous systems will be heavily biased in their responses, and new users will require very long training to develop the tricks and heuristics for the normal task performance which will be necessary for an empirical test of the functional fixations during unfamiliar situations. The question raised in the present paper is whetner empirical evaluation is feasible at all for the kind of systems considered, and/or whether further development of a framework for analytical evaluation is necessary.

SYSTEM EVALUATION AND RISK ANALYSIS

The difficulties in evaluating new designs for application in which the performance during infrequent, unfamiliar task conditions is of prime importance, are related to the fact that human performance cannot be decomposed into elements or functions which can be studied experimentally in isolation, as is the case for technical systems. Human response to cues from the environment very much depends on the context: trained operators compose their responses from a repertoire of skilled subroutines, they are preconditioned for responses by their process feel, i.e. their expectations, and responses are frequently released by simple and informal cues. The only feasible approach to evaluation of a system at the level of effectiveness

for support during unfamiliar task conditions seems to be an analytical evaluation based on a conceptual model of man-machine interaction. This model should represent the structure of the human cognitive control in such a way that it makes it possible to assess the likely interaction between more elementary human functions, in particular the biases posed on infrequent responses from the background of habitual routines and familiar perceptual patterns. The basis of such an assessment will be information on more elementary patterns of behaviour studied separately, either during real life performance, in part task simulators, or by dedicated psychological experiments.

The problems met are analoguous to those of probabilistic risk assessment. Here, the probability of certain rare events has to be judged by means of a model representing the causal interaction among components during the specific scenario, from failure probabilities collected for the components individually. In man-machine system evaluation, the human behaviour during unfamiliar situations will be judged from information on more elementary behaviour by means of a model of the interaction of such elements, not only during a specific scenario, but across tasks and time. Another difference is that probabilistic risk assessment is based on prediction of **particular courses of events** aggregated in families (event trees), whereas due to the adaptive nature of human performance man-machine system evaluation has to be based on relationships among categories of tasks and functions to allow for individual differences. Rouse et al., for instance, base their approach on generic decision functions, "prototype messages", and frequency classification of tasks (Rouse et al., 1984).

A basic similarity is that neither probabilistic risk assessment nor system evaluation can be considered as a well-formed, isolated activity. Probabilistic risk analysis is a computational exercise relating a model, a set of assumptions, and some data to a global risk figure. This figure has no meaning unless the model and computational assumptions are considered as specifications for the ultimate acceptable operating conditions, and the correspondence is verified during operation. The risk assessment is not a separate analysis for acceptance of the design, it is the explicit formulation of the safety properties of the system to serve as a coupling of the design basis and the operations planning and management (Rasmussen and Pedersen, 1984).

In the same way it will be important to consider an analytical verification of a design, not only as a test of the internal

consistency of a design and of the correspondence with the final implementation, but basically as the explicit formulation of the design basis which will make it possible to plan system application, work organisation, and training of the end-users. Similarly, the validation is less an analysis of the design in order to evaluate the match between the system and the real life conditions, than it is an analysis of the future conditions of use, to see whether it will be realistic to match these conditions to the assumptions behind the design. In case of a less satisfying match, modifications of either side may be feasible. Rouse et al. already stressed the importance of a coordinated design and evaluation; this should be extended to a consideration of the evaluation analysis to be the formalisation of the design basis to serve as the link between design and operation planning. Furthermore, this should be a dynamic process which continues during the total phase of operation; no organisation is stable, neither is the level of training of users, and a dynamic interaction between operations experience and properties of a decision support system may be crucial. Just as the value of a probabilistic risk analysis depends on a coordination with an effective risk management, which serves to upgrade both the system and the operating practice according to the operational experience gained from the analysis of event reports. In this way, the in-plant evaluation which is the last step in the evaluation procedure proposed by Rouse et al. should be a continuous process.

STRUCTURE OF A CONCEPTUAL FRAMEWORK

From this discussion it follows that conceptual models have two different roles in system evaluation. One is to serve as a vehicle for the explicit formulation of the design concept which is necessary as a reference for operation planning and for evaluation. For the more frequent task situations this evaluation can be performed empirically. Another role is to serve as an analytical evaluation for those rare situations which cannot be studied empirically during design, and for which human performance has to be predicted by means of a conceptual model interrelating more basic knowledge of human behaviour generalised from a wide variety of sources. The basic structure of such a conceptual framework will include several components: a systematic representation of the task domain, i.e. of the problem space in which decision making will have to take place

during unfamiliar situations; a set of normative models of the information processes which are acceptable for the task from a system effectiveness point of view, i.e. the information processes which can be used; and, finally, a model representing performance of humans considered as system components. This model will have to include several aspects. Models which describe human resources and limitations with respect to the normative task formulations are necessary, in order to judge the design of interfaces and training schemes. Other models at a higher level are necessary to judge the criteria behind the subjective task formulation and the actual choice of information processing strategy, as well as the interaction between more elementary behavioural elements in order to judge which information processes will be used.

The elements in such a framework (Rasmussen, 1984, 1985) will be discussed in some more detail in the following sections.

The problem domain. The first aspect to consider will be the problem domain, i.e. the relationships in the many-to-many mappings of the purpose-function-process-equipment hierarchy. These means-end relationships should be analysed and systematically described along the part-whole and means-end dimensions in order to have a consistent framework for identification of the control requirements of the system and the content of the related decision tasks. Supervisory disturbance control is a resource management task in the means-end hierarchy, and adequacy of decision support cannot be judged without an explicit description of the system in terms of the configuration and state of the available resources at each level.

This is an engineering analysis, and will be a top-down explicitation of the bases for the design decisions in technical terms. As mentioned above, this analysis is particularly important during the introduction of new technology for control systems and man-machine interfaces, in order to replace the industrial practice implicit in the plant design by explicit models as a reference for evaluation of those functional purposes and constraints which are considered crucial for the judgement of plant performance. The need for such an analysis was realised during the feasibility studies for disturbance analysis systems following the Three Mile Island incident (Gallagher et al., 1982), and an attempt to formalise the description has been made by Lind (1982). An application on existing systems has shown that an afterrationalisation of an existing plant design in order to get a rational basis for a new control system design involves a considerable amount of work.

The decision task to be performed in this problem domain can be characterised in two respects; one is a description of the elements in the decision task, another is the information processing strategies which are applicable within these elements. The decision sequence required for typical situations can be described in terms of the elements of analysis and diagnosis; evaluation and goal prioritising; and planning and execution. For proper evaluation of a system it is very important that the emergency management philosophy is made explicit, i.e. the criteria behind the decisions, how these functions should be shared by the operator, the computer, and the systems designer. The evaluation will have to consider whether the decision should be based on the designer's a priori analysis, which is therefore implemented in terms of operational instructions and/or computer programs, or whether the decision must be left for an on-line evaluation by operators and/or computers. The task allocation adopted for a particular scenario will influence the appropriate design of interfaces, staffing, and training schemes, and a description will be necessary as a reference for any attempt to evaluate the design. Just choosing subjects for empirical evaluation with a training or practical experience which does not match the allocation assumed will lead to erroneous judgements. Frequently, in particular in systems based on more traditional technology, this information is only implicitly available, and a judgement and classification of the particular system by the evaluator as suggested by Rouse et al. will be necessary to identify a reference for judgement. The ambiguity has been demonstrated by the development in emergency management philosophy after Three Mile Island. Two different lines of development have been discussed. One being a further formalisation in direction of "symptom-based" procedures for which also the diagnostic phase is preplanned by the designer, the other being the call for more "knowledge-based" operator decisions. In both cases the operators are supposed to follow certain rule sets supported by training sessions and to monitor the performance by intelligent evaluation, and the designer will (more or less implicitly) choose to support these functions in varying degrees depending upon his perception of management policy for allocation of authority to operators, designer, and automatic equipment. In this situation, the reliability of an evaluation very much depends on the possibility of judging the correspondence between assumptions behind design and behind operation management.

Strategies and heuristics. Other aspects of the mental task to consider are the information processing strategies and heuris-

tics which will be effective and acceptable for the different decision functions, and which may be used on-line by the computer or by an operator. Some conception of such strategies will be used as design basis for the display formats and for planning the content of training programmes. Different strategies for the same mental task will have very different requirements with respect to processing capacity, amount of observation and nature of background knowledge, and explicit formulation of the design basis is crucial for proper coupling to evaluation as well as to operator training. The conceptual models of the information processing strategies needed for evaluation will not be detailed process models, but higher level models of prototypical strategies in terms of the structure and content of the mental model required, the amount and type of data on the actual system states and reference states, and generic strategical rules. Such information will enable an identification of the content in prototypical messages from the interface and the structure of the display format which is required for an evaluation of effectiveness of a decision support. Such an evaluation will have to consider whether the actual design offers an envelope around a strategy which will be effective for the scenario considered, and within which an operator can improvise an acceptable decision process.

The human information processor. The conceptual framework discussed so far will make it possible to judge analytically whether the supervisory control decisions required in a scenario adapted for evaluation can be made, i.e. whether the information needed for an effective decision process will be presented. An evaluation remains whether the proper decision will be made, i.e. whether the decision maker will have the adequate resources and choose the proper strategy in the actual situation, and perform it successfully.

The first of these questions, whether operator performance will be inadequate due to data or resource limitations, can be approached for the different possible decision strategies and scenarios separately by a combination of analytical evaluations based on data from psychological research and empirical evaluations by simulations. The hard bit is the second question, whether operators in the actual situation will use the proper subjective formulation of the task, and choose to use a proper strategy. This cannot be empirically resolved. Systematic experiments with rare events are impossible by definition, and operators having the proper state of training with a new system concept are not available at the proper time. Analytical assessment therefore has to be used as far as it can bring one.

However, since "the whole of a man-machine system is more than the sum of its parts" (Hollnagel and Woods, 1983), this evaluation has to be based on a model of human performance at a higher level than the usual psychological models, since it has to represent the overall cognitive control of behaviour in a complex real life situation. It has to describe the interaction of various human functions which are related to different stages of training, such as automated sensori-motor skills, use of heuristic rules, and problem solving. These are functions which are normally studied separately in psychological research. Furthermore, the models should represent the limitations of human capabilities, and explicitly consider those psychological mechanisms which can lead to errors.

A first step to such a model has been the skill-rule-knowledge model of cognitive control which has been used to characterise human errors (Rasmussen, 1980). Recently this framework has been the basis of an analysis of human errors which further supports the view that a large fraction of human slips and mistakes can be explained by a few cognitive mechanisms which are related to the interference in unfamiliar situations from patterns learned during routine tasks (Reason, 1985). Reason discusses several types of human errors and decision biases with reference to a higher level model of cognitive control mechanisms, and the approach seems to be useful for development of guides for analytical evaluation of the form and content of messages from a decision support system.

It is, however, not only a question of the reliability of the global psychological models. The dependence of performance during infrequent circumstances upon habits and routines requires for evaluation a description not only of the particular emergency scenario, but of the total task repertoire. This should be done in a way which makes it possible to evaluate the "Hamming distance" between the symptom set related to the situation considered and those related to more frequent routine tasks. The tool for such an analysis could be the "confusion matrix" which makes it possible to identify the likely false associations to familiar procedures which may happen if one or more cues in a particular symptom set are overlooked. This kind of analysis is becoming part of risk analysis, and a coordination of risk analysis and interface evaluation appears to be beneficial. In addition, the scenarios developed for operator response tree analysis in risk analysis are well suited to serve as context for the evaluation of the man-machine interface (for a recent review of the state of cognitive models for risk analysis, see Hannaman, 1985).

CONCLUSION

This analysis only touches part of the evaluation problem. It is implicitly assumed that an operator will accept the support of a computer in his supervisory decision task, and that he will trust the information and advice received from it. This may, however, not be the case, and criteria for judging user acceptance and role allocation in systems aiming at interactive decision making are badly needed. Under which conditions will a human user understand the performance of a computer in a decision task and accept its advice or instructions during a complex and possibly risky plant condition?

There are still many problems to be solved before a complete analytical evaluation is possible. This should, however, not prevent the use of the presently available knowledge in a systematic design and evaluation and a proper coordination to the operational planning and operator training.

REFERENCES

Gallagher, J.M., Easter, J.R., Rumancik, J.A., Campbell, L.A. and Mueller, N.P. (1982): Disturbance Analysis and Surveillance System Scoping and Feasibility Study. Report NP-2240 Electrical Power Research Institute, Palo Alto, Ca.

Hannamann, G.W., Spurgin, A.J. and Lukic, Y.D. (1984): Human Cognitive Model for PRA Analysis. Report NUS-4531 NUS Corporation San Diego, Ca. For Electrical Power Research Institute. In Press.

Hollnagel, E. (1981): Verification and Validation in the Experimental Evaluation of New Designs for Man-Machine Systems. Risø-M-2300.

Hollnagel, E. and Woods, D. (1983): Cognitive Systems Engineering: New Wine in New Bottles. Int. J. of Man-Machine Studies, 18, p. 583-600.

Kuhn, T.S. (1962): The Structure of Scientific Revolution. Univ. of Chicago Press. Chicago Ill.

Lind, M. (1982): Multilevel Flow Modelling of Process Plants for Diagnosis and Control. International Meeting on Thermal Nuclear Reactor Safety, August 1982, Chicago, Ill.

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): Strategies for State Identification and Diagnosis in Supervisory Control Tasks, and Design of Computer-Based Support Systems. In: W.B. Rouse (Ed): Advances in Man-Machine Systems Research, Vol.1. J.A.I. Press Inc.

Rasmussen, J. (1985): On Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. Elsevier Scientific Publishing Co.

Rasmussen, J. and Pedersen O.M. (1984): Human Factors in Probabilistic Risk Analysis and in Risk Management. In: Operational Safety of Nuclear Power Plants, Vol. 1, I.A.E.A., Vienna.

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.

Rouse, W.B. (1984): Computer-Generated Display System Guidelines: Volume 2, Developing an Evaluation Plan. NP-3701, Vol 2. Search Technology, Atlanta under ORNL contract No. W-7405-eng-26 for Electrical Power Research Institute Palo Alto, Ca.

Rouse, W.B., Frey, P.R. and Rouse, S.H. (1984): Classification and Evaluation of Decision Aids for Nuclear Power Plant Operators. Report no. 8303-1. Search Technology, Atlanta under ORNL contract No. 62X-43185V for Nuclear Regulatory Commission, Washington DC.

Woods, D.D., Wise, J., and Hanes, L. (1981): An Evaluation of Nuclear Power Plant Safety Parameter Display Systems. In the Proceedings of the 24th Annual Meeting of the Human Factors Society. Santa Monica, CA.

Risø National Laboratory **Risø - M -** 2520

Risø - M - 2520

Title and author(s)

Conceptual Models in Man-Machine Design Verification

Jens Rasmussen

Date
August 1985

Department or group
Electronics

Group's own registration number(s)
R-6-85

pages + tables + illustrations

Abstract

The need for systematic methods for evaluation of design concepts for new man-machine systems has been rapidly increasing in consequence of the introduction of modern information technology. Direct empirical methods are difficult to apply when functions during rare conditions and support of operator decisions during emergencies are to be evaluated. In this paper, the problems of analytical evaluations based on conceptual models of the man-machine interaction are discussed, and the relations to system design and analytical risk assessment are considered. Finally, a conceptual framework for analytical evaluation is proposed, including several domains of description: 1. The problem space, in the form of a means-end hierarchy; 2. The structure of the decision process; 3. The mental strategies and heuristics used by operators; 4. The levels of cognitive control and the mechanisms related to human errors. Finally, the need for models representing operator's subjective criteria for choosing among available mental strategies and for accepting advice from intelligent interfaces is discussed.

Copies to

Available on request from Risø Library, Risø National Laboratory (Risø Bibliotek), Forsøgsanlæg Risø), DK-4000 Roskilde, Denmark
Telephone: (03) 37 12 12, ext. 2262. Telex: 43116