The Role of Error in Organizing Behaviour

Rasmussen, Jens

Publication date:
1989

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

Risø-M-2799

RISØ

Risø-M-2799

The Role of Error in Organizing Behaviour

Jens Rasmussen

Risø National Laboratory, DK-4000 Roskilde, Denmark
July 1989

Risø National Laboratory

Risø – M – 2799

Title and author(s)

THE ROLE OF ERROR IN ORGANIZING BEHAVIOUR

Jens Rasmussen

Date July 1989

Department or group

Information Technology

Groups own registration number(s)

R-2-89

Project/contract no.

Pages 18 Tables Illustrations 7 References 14

ISBN 87-550-1543-3

Abstract (Max. 2000 char.)

ABSTRACT

The paper reviews the significance of the concept of human error in three different applications: task analysis and human reliability estimation, causal analysis of accidents, and design of reliable socio-technical systems. It is concluded that the definition of error is ambiguous and depends on the purpose of analysis. For analysis of accidents, the causes identified depend on the stop-rule applied to terminate the analytical backtracking, and the stop-rule will be different depending on whether the purpose of the analysis is the explain the case, to identify a responsible person, or to improve safety. It is concluded that it can be difficult by a linear causal analysis to identify structural properties leading to such systemic failures which can be caused by adaptive features of the system.

Finally, it is shown how adaptation guided by subjective process criteria serves to resolve the degrees of freedom in a work environment so as to remove the need for choice and decision in a stable environment. In a modern, varying environment, however, need for adaptation remains and the inevitable effect of adaptation will be various forms of error. Reliability and safety of socio-technical systems, consequently, depend on boundaries around acceptable performance that are reversible and allow recovery in case of violation. Improvement of safety depends on graceful loss of control and opportunity for learning to recover, rather than on withdrawal of boundaries from present normal operation.

Descriptors

Available on request from Risø Library, Risø National Laboratory, (Risø Bibliotek, Forskningscenter Risø),
P.O. Box 49, DK-4000 Roskilde, Denmark.
Telephone 02 37 12 12, ext. 2262. Telex: 43116, Telefax: 02 36 06 09

RR'950000 99

RISØ-M-2799

THE ROLE OF ERROR IN ORGANIZING BEHAVIOUR

Jens Rasmussen

ABSTRACT

The paper reviews the significance of the concept of human error in three different applications: task analysis and human reliability estimation, causal analysis of accidents, and design of reliable socio-technical systems. It is concluded that the definition of error is ambiguous and depends on the purpose of analysis. For analysis of accidents, the causes identified depend on the stop-rule applied to terminate the analytical backtracking, and the stop-rule will be different depending on whether the purpose of the analysis is the explain the case, to identify a responsible person, or to improve safety. It is concluded that it can be difficult by a linear causal analysis to identify structural properties leading to such systemic failures which can be caused by adaptive features of the system.

Finally, it is shown how adaptation guided by subjective process criteria serves to resolve the degrees of freedom in a work environment so as to remove the need for choice and decision in a stable environment. In a modern, varying environment, however, need for adaptation remains and the inevitable effect of adaptation will be various forms of error. Reliability and safety of socio-technical systems, consequently, depend on boundaries around acceptable performance that are reversible and allow recovery in case of violation. Improvement of safety depends on graceful loss of control and opportunity for learning to recover, rather than on withdrawal of boundaries from present normal operation.

July 1989

Risø National Laboratory

Invited Contribution to the CEC Workshop on Errors in Operation of Transport Systems;
MRC-Applied Psychology Unit, Cambridge May 1989.

ISBN 87-550-1543-3
ISSN 0418-6435

Grafisk Service, Risø 1989

CONTENTS Page

Introduction

During recent years, the significance of the concept of human error has changed considerably. The reason for this has partly been an increasing interest of psychological research in analysis of complex real-life phenomena, partly the changes of modern work conditions caused by advanced information technology. Consequently, the topic of the present contribution will not be a definition of the concept or a proper taxonomy. Instead, a review will be given of a couple of professional contexts for which the concept of error is important.Three cases of analysis of human-system interaction will be reviewed: 1. Traditional task analysis and human reliability estimation; 2. Causal analysis of accidents after the fact and finally, 3. Design of reliable work conditions in modern socio-technical systems. It is concluded that 'errors' cannot be studied as a separate category of behaviour fragments; the object of study should be cognitive control of behaviour in complex environments.

1. Traditional task analysis and human reliability estimation

Human activity in traditional work environments can be described in terms of repetitive tasks, i.e., sequences of acts in control of some equipment or tool. Manufacturing systems were normally planned for effective and economic operation during long periods of time. Planned or normal work sequences had time to settle in stable patterns which could be identified during design by analysis of the task to control tools and equipment or afterwards by field studies. Since successful operation during production or in a particular mission was of fundamental interest, technical and human reliability analysis became important design tools both for military and high hazard industrial operations.

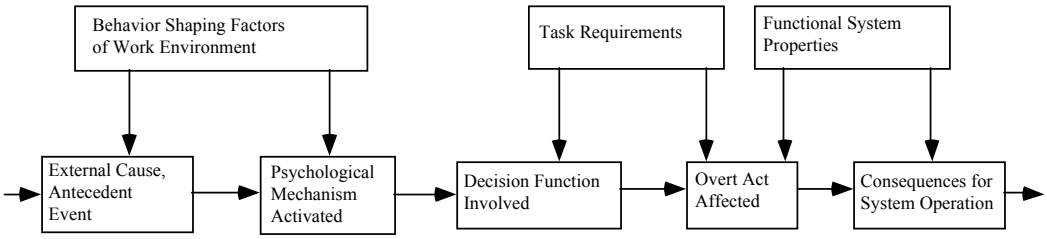

Figure 1 illustrates the human involvement in a causal sequence of events. The event of human error is decomposed to identify the cognitive task element and the psychological mechanism involved in the error. At this level of detail, an event in the work context activates a particular psychological mechanism which influences the immediate decision task required by the work. A decision error in turn, introduces an error in the overt action sequence with unacceptable consequences for the work goal. Two aspects are essential in the present context. One is that human 'error' very frequently will be a link in a sequence, not the origin. Secondly 'errors' can be categorized at different stages in the flow. This representation is well suited for Failure-Mode-And-Effect analysis in the interface of technical systems: The various psychological 'error mechanisms' are folded onto the cognitive task from which the effects, in turn, are folded onto functional system properties for evaluation of the consequences.

In this situation, human errors can easily be defined; normative sequences of proper acts are available for reference and errors can readily be identified and recorded. As long as analysis is concerned with familiar, repetitive tasks, errors caused by lack of resources or proper intention are of minor importance and errors can be studied in terms of their overt effects (Swain's Therp method, Swain and Guttmann, 1983). In modern work places people are frequently moved to supervisory tasks and decision making and, consequently, reliability analysis will be focused on less well-structured and stable tasks

involving diagnosis and contingency planning. Focus of error analysis is moved back from overt acts to decision functions and further on to psychological mechanisms (see figure 1). It is a remarkable fact that given a particular sequence of human acts, taxonomies of error analysis resulting from detailed analysis of actual cases of incidents and accidents and from psychological laboratory research show definitive convergent properties (Rasmussen, 1988a). When a particular task sequence can be taken as reference (i.e., a sequence which is functionally constrained by the equipment to operate or firmly established by training) a failure-mode-and-effect analysis is a very feasible approach to identify the hazards presented by human error. It will be effective during design to ensure error tolerance even if quantitative reliability prediction may not be realistic (Rasmussen, 1982).

A necessary precondition is, however, that the sequence in which the 'error' is analyzed can be taken for granted. This is the case only when we are involved in a *local* analysis focused on the immediate human-machine interface: We then try to predict the risk involved in the operation of some particular technological system of a known design. The acceptable work procedure is identified from the functional requirements of equipment, given a definite goal. This is, as mentioned, a reasonable assumption if the task is repetitive as it was the normal case in established technology. In addition, we are dealing with a human link in an extended chain of events; the 'error' is a link in the chain, in most cases *not* the origin of the course of events. This kind of analysis and consequently, definition of error is completely inadequate when we are dealing with design or improvement of large-scale socio-technical systems. In general, we do not have a simple causal trace deflected from its intended course toward one goal. Actually, such a separate trace is the manifestation of a particular, dynamic flow of events in a complex network involving several goals and side effects and many side branches. Previous flows of events along these branches have served to precondition the river bed in which the dynamic flow is taking place.

2. Causal analysis of accidents after the fact

In this case, we are analyzing a chain of events up-stream from an accident in order to understand, *why* it happened; to find somebody to blame, *who* done it; or find out *how* to improve the system. We are trying to describe a particular course of events and to identify the particular causal trace in which human error is embedded.

Accidents are normally analyzed in terms of accidental chains of events, i.e., causal representations. Since no two accidents will be identical, accident analysis will depend on prototypical categories of causes, events, and consequences. Representation of the behaviour of the physical world in causal terms is very effective for describing accidents because the objects of the real world are explicit in the model and their changes are easily modelled which is not the case in a model based on relations among quantitative variables. It is, however, important to consider the implicit frame of reference of a causal analysis (Rasmussen, 1988b).

The behaviour of the complex, real world is a continuous, dynamic flow which can only be explained in causal terms after decomposition into discrete events. The concept of a causal interaction of events and objects depends on a categorisation of human observations and experiences. Perception of occurrences as events in causal connection does not depend on categories which are defined by lists of objective attributes but on categories which are identified by typical examples, i.e., prototypes (as defined by Rosch, 1975). This is the case for objects as well as for events. Everybody knows perfectly well

what 'a cup' is. To define it objectively by a list of attributes that separates cups from jars, vases and bowls is no trivial problem. It has for instance, been met in many attempts to design computer programs for picture analysis. The basic problem is that the property to be 'a cup' is not a feature of an isolated object but depends on the context of human needs and experience. The identification of events in the same way depends on the relationship in which they appear in a causal statement. An objective definition, therefore, will be circular.

In the analysis of accidents, decomposition of the dynamic flow of changes will normally terminate when a sequence is found including events which match the prototypes familiar to the analyst. The resulting explanation will take for granted his frame of reference and in general, only what he finds to be unusual will be included: the less familiar the context, the more detailed the decomposition. By means of the analysis, a causal path is found up-stream from the accidental effect. This path will be prepared by resident conditions which are latent effects of earlier events or acts. Also these resident conditions can be explained by causal back-tracking and in this case branches in the path are found. To explain the accident, these branches are also traced backward until all conditions are explained by abnormal, but familiar events or acts. The point is: how do the degree of decomposition of the causal explanation and selection of the side-branches depend on the circumstances of the *analysis*? Another question is: What is the stop-rule applied for termination of the search for causes? Ambiguous and implicit stop rules will make the results of analysis very sensitive to the topics discussed in the professional community at any given time. There is a tendency to see what you expect; during one period, technical faults were in focus as causes of accidents, then human errors predominated while in the future focus will probably move up-stream to designers and managers. This points to the question whether system break-down is related to higher level functional structures and feedback mechanisms rather than to the local conditions of events. In that case, traditional causal attribution turn out to be fighting symptoms rather than the origin of break-down.

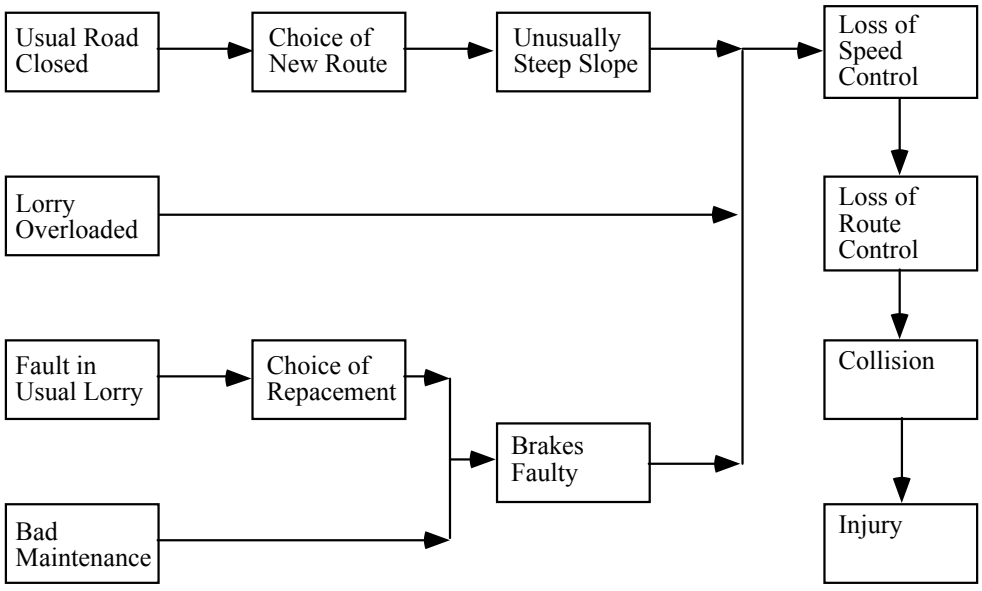

Figure 2 illustrates a causal explanation of a driving accident. The flow of behaviour is decomposed into chains of events. Note that only abnormal or unusual events together with violations of rules are included. The normal activities conditioning the path are not included. Furthermore, decomposition and causal backtracking stop at events which are taken to be 'reasonable explanations.'

(Adopted from Leplat & Rasmussen, 1987).

The perception of stop-rules is very important in the control of causal explanations. Everyone from college knows the relief felt when finding a list of solutions to math problems. Not that it gave the path to solution to any great extent, but it gave a clear stop-rule for the search for possible mistakes, overseen preconditions, and calculation errors. The result: hours saved and peace of mind. A more professional example to the same point is given by Kuhn (1976). He mentions the fact that chemical research only was able to come up with whole-number relations between elements of chemical substances after the acceptance of John Dalton's chemical atom theory. There had been no stop rule for the efforts in refinement of the experimental technique until the acceptance of this theory.

Stop-rules are not usually formulated explicitly. The search will typically be terminated pragmatically in one of the following ways: (a) An event will be accepted as a cause and the search terminated if the causal path can no longer be followed because information is missing; (b) A familiar, abnormal event is found to be a reasonable explanation; or (c) A cure is available. The dependence of the stop rule upon familiarity and the availability of a cure makes the judgement very dependent upon the role in which a judge finds himself. An operator, a supervisor, a designer, and a legal judge will reach different conclusions.

To summarize: identification of accident causes is controlled by pragmatic, subjective stop-rules. These rules depend on the aim of the analysis, i.e., whether the aim is to explain the course of events, to allocate responsibility and blame, or to identify possible system improvements in order to avoid future accidents.

Analysis for Explanation. In an analysis to explain an accident, the backtracking will be continued until a cause is found *which is familiar* to the analysts. If a technical component fails, a component fault will only be accepted as the prime cause if the failure of the particular type of component appears to be 'as usual.' Further search will probably be made, if the consequences of the fault make the designer's choice of component quality unreasonable, or if a reasonable operator could have terminated the effect, had he been more alert or been better trained. In such a case, a design or manufacturing error, respectively an operator error will be accepted for explanation.

In most recent reviews of larger industrial accidents, it has been found that human errors are playing an important role in the course of events. Frequently, *errors are attributed to operators involved in the dynamic flow of events*. This can be an effect of the very nature of the causal explanation. Human error is, particularly at present, familiar to an analyst: to err is human, and highly skilled people will frequently depart from normative procedures as we will see in a subsequent section.

Analysis for Allocation of Responsibility. In order to allocate responsibility, the stop-rule of the backward tracing of events will be to identify a person who made an error and at the same time, '*was in power of control'* of his acts. The very nature of the causal explanation will focus attention on people directly and dynamically involved in the flow of abnormal events. This is unfortunate because they can very well be in a situation where they do not have the 'power of control.' Traditionally, a person is not considered in power of control when physically forced by another person or when subject to disorders such as e.g., epileptic attacks. In such cases, acts are involuntary (Fitzgerald, 1961; Feinberg, 1965), from a judgement based on physical or physiological factors. It

is, however, a question as to whether cognitive, psychological factors should also be taken more into account when judging 'power of control.' Inadequate response of operators to unfamiliar events depends very much on the conditioning taking place during normal work. This problem also raises the question of the nature of human error. The behaviour of operators is conditioned by the conscious decisions made by work planners or managers. They will very likely be more 'in power of control' than an operator in the dynamic flow of events. However, their decisions may not be considered during a causal analysis after an accident because they are 'normal events' which are not usually represented in an accident analysis. Furthermore, they can be missed in analysis because they are to be found in a conditioning side branch of the causal tree, not in the path involved in the dynamic flow.

Present technological development toward high hazard systems requires a very careful consideration by designers of the effects of 'human errors' which are commonplace in normal, daily activities, but unacceptable in large-scale systems. There is considerable danger that systematic traps can be arranged for people in the dynamic course of events. The present concept of 'power of control' should be reconsidered from a cognitive point of view, as should the ambiguity of stop-rules in causal analysis to avoid unfair causal attribution to the people involved in the dynamic chain of events.

*Analysis for System Improvements***.** Analysis for therapeutic purpose, i.e., for system improvement, will require a different focus with respect to selection of the causal network and of the stop-rule. The stop-rule will now be related to the question of whether an effective *cure is known*. Frequently, cure will be associated with events perceived to be 'root causes'. In general, however, the effects of accidental courses of events can be avoided by breaking or blocking any link in the causal tree or its conditioning side branches. Explanatory descriptions of accidents are, as mentioned, focused on the unusual events. However, the path can also be broken by changing normal events and functions involved. The decomposition of the flow of events, therefore, should not focus on unusual events, but also include normal activities.

The aim is to find conditions sensitive to improvements. Improvements imply that some person in the system makes decisions differently in the future. How do we systematically identify persons and decisions in a (normal) situation when it would be psychologically feasible to ask for a change in behaviour as long as reports from accidents focus only on the flow of unusual events? An approach to such an analysis for improving work safety has been discussed elsewhere (Leplat and Rasmussen, 1984).

In conclusion, the choice of stop-rules for the analysis of accidents is normally left to the subjective judgement of the analyst, depending heavily on the aim of his analysis. Analyses made for one purpose may, therefore, be misleading for other purposes.

3. Design of reliable work conditions and socio-technical systems

Modern Work Conditions. A number of problems are met when attempts are made to improve safety of socio-technical systems from analyses tied to particular paths of accidental events. This is due to the fact that each path is a particular token shaped by higher order relational structures. If changes are introduced to remove the conditions of a particular link in the chain, odds are that this particular situation will never occur again. We should be fighting types, not individual tokens. Human behaviour is constrained in a way that makes the chain of events reasonably predictable only in the immediate interface to the technical systems. The longer away from the technical core

we are, the more degrees of freedom agents have in their mode of behaviour. Consequently, the less certain is also the reference in terms of normal or proper behaviour for judging 'errors'. This problem is becoming increasingly important as the modern manufacturing systems and organizations are forced to become highly flexible in order to be able to respond to increasingly dynamic market requirements, technological innovations, and legal constraints.

In this situation, improvements of safety features of a socio-technical system depend on a *global* analysis: No longer can we assume the trace of human behaviour to be predictable. Tasks will be formed for the occasion, and design for improvements must be based on attempts to find means of control at higher levels than the level of particular task procedures. If, for instance, socio-technical systems have features of adaptation and self-organization, changes to improve safety at the individual task level can very well be compared to attempts to control the temperature in a room with a thermostat-controlled heater by opening the window. In other words, it is not sensible to try to change performance of a feedback system by changes inside the loop, you have to identify mechanisms that are sensitive, i.e., related to the control reference itself.

Such basic, high level features of "human error" in flexible socio-technical system are related to the dependence of human performance on features such as 1. Learning and adaptation, 2. Conflicts among cognitive control structures, 3. Resource limitations and finally, 4. Stochastic variability. An attempt to develop guidelines for design of human-work interfaces from such higher level features has been presented elsewhere (Rasmussen and Vicente, 1987).

Human adaptation. In all work situations constraints are found which must be respected to perform satisfactorily. There are, however, also many degrees of freedom which have to be resolved at the worker's discretion. In stable work conditions, know-how will develop which represents prior decisions and choice and the perceived degrees of freedom will ultimately be very limited, i.e., 'normal ways' of doing things will emerge, and the process of adaptation will no longer be messing-up the concept of error. In contrast, in modern, flexible and dynamic work conditions, the immediate degrees of freedom will have to be continuously resolved. This implies that effective work performance includes continuous awareness of the available degrees of freedom together with effective strategies for making choice, ahead of the task of controlling the chosen path to a goal. This changes the concept of error in a very fundamental way.

The behaviour in work of individuals (and, consequently, also of organizations) is, by definition, oriented towards the requirements of the work environment as perceived by the individual. Work requirements, *what* should be done, will normally be perceived in terms of control of the state of affairs in the work environment according to a goal, i.e., *why* it should be done. *How* these changes are made to a certain degree is a matter of discretion of the agent.

The *alternative, acceptable work activities*, how to work, will be shaped by the work environment which defines the boundaries of the space of *possibilities*, i.e., acceptable work strategies. This space of possibilities will be further bounded by the resource profile of the particular agent in terms of tools available, knowledge (competence), information about state of affairs, and processing capacity. The presence of alternatives for action depends on a many-to-many mapping between means and ends present in the work situation as perceived by the individual; in general, several functions can serve the individual goals and each of the functions can be implemented by different tools and

physical processes. If this was not the case, the work environment would be totally predetermined and there would be no need for human choice or decision.

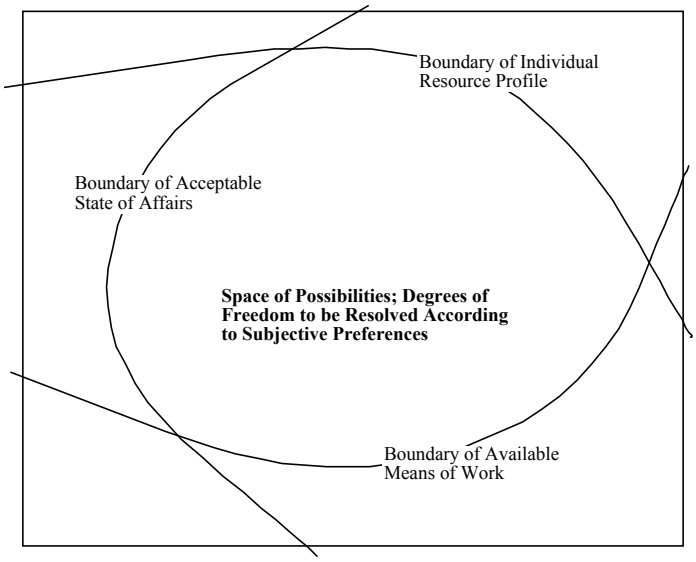

Figure 3. Human behaviour is governed by constraints which must be respected by the actors for work performance to be successful. Identification of such constraints will specify the 'space' in which the human can navigate freely. Violation of the constraints will be considered human error or task violation in the usual sense. For successful performance, humans have to navigate between several boundaries. One is given by the control requirements posed by the system, another by the means offered for work. □et another boundary is given by the human resource profile which depend on individual characteristics such as competence, mental capacity, physical strength, etc. Navigation within the envelope specified by these boundaries will depend on subjective criteria for choice, such as aim to save time, to spare memory load, to have fun, to explore new land, etc.

Within the space of acceptable work performance between the boundaries defined by the work requirements on one side and the individual resource profile on the other side, considerable degrees of freedom are still left for the individual to choose among strategies and to implement them in particular sequences of behaviour. These degrees of freedom must be eliminated by the choice of an agent to finally enter a particular course of action. The different ways to accomplish work can be categorized in terms of *strategies*, defined as *types* of behavioural sequences which are similar in some well defined aspects, such as the physical process applied in work and the related tools or, for mental strategies, the underlying kind of mental representation and the level of interpretation of perceived information. In actual performance, a particular situation-dependent exemplar of performance, a *token* will emerge which is an implementation of the chosen strategy under the influence of the complexity of detail in the environment. The particular token of performance will be unique, impossible to predict, whereas the strategy chosen will, in principle, be predictable. This choice made by the individual agents depends on *subjective performance criteria* related to the process of work such as time spent, cognitive strain, joy, cost of failure, etc. In general the freedom to chose work strategy and to shift dynamically between strategies is very important as a means to resolve resource-demand conflicts met during performance.

Figure 4. An Example: The activities involved in going to work. The *work given constraints* are related to the location, the time of arrival, and the probability of delays. *Constraints in means* are defined by the transport alternatives, i.e., to take the tube, a taxi, or to drive by yourself. The *subjective process criterion* determining your choice depend on economy, your wife´s request to bring some grocery and maybe, consideration of the time spent, the likelihood of traffic jams. Given the decision to drive by yourself, the choice of route depends on the secondary task of shopping, of your joy with a particular scenery, and the traffic density. Finally, en route, the speed you choose depend on traffic given constraints, on

formal conditions such as speed limits or your wife´s anxiety, and ultimately sporty criteria related to your driving skill, i.e., to drive fast and smoothly

	Means-Ends Relations
Goals and Values, Constraints	Work, Income, Intellectual and Esthetic Pleasure; Social Family Relations
Priority Measures, Flow of money, material, Information	Expenses, Joy, probability of Success and Delays
General Functions	Work Function: Transport Family Function: Shopping
Physical Processes in Work and Equipment	Train Operation, Schedules, Space for Reading; Taxi: Time Spent, Traffic Jams, Price; Car: Traffic, Time, Price
Material Objects, Tools, Buildings, etc. Topography	Train, Taxi, Private Car Routes, Distances, Scenery; Shops and Locations, etc.

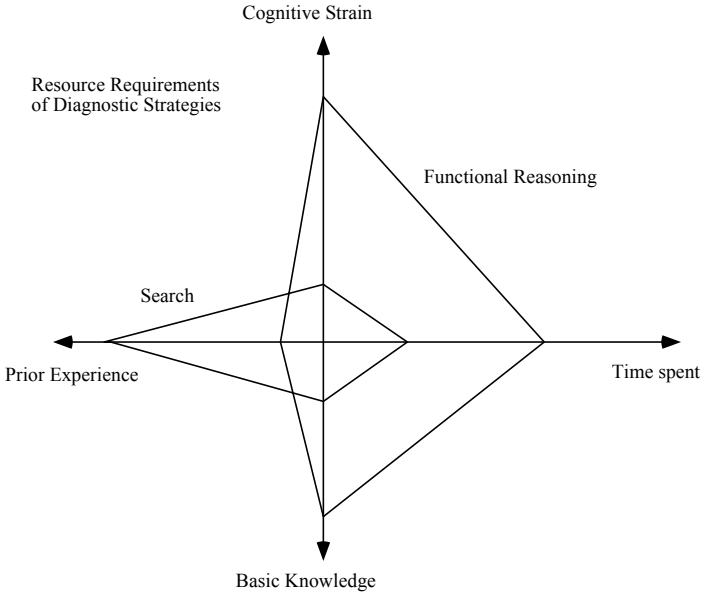

Figure 5. The figure serves to illustrate the different resource re-quirements of different mental strategies. This difference makes a shift in strategy when faced with difficulties in a task, an effective way to navigate along the path of least effort, a very popular strategy in skilled performance to adapt behaviour to immediate work situation.

Modelling work activity from this point of view depends on identification of the space of acceptable and possible work strategies (i.e., prototypical sets of behaviour sequences), the human resource profile, and the subjective criteria governing the resolution of the remaining degrees of freedom in different work scenarios. Some work requirements are explicit and discrete with specified limits of acceptance. Other requirements are formulated as optimising criteria serving to resolve ambiguity in goal specification, such as the request to reach a solution which is as cheap or as safe as possible. Such *product* criteria, together with the subjective *process* criteria, necessarily will lead to an adaptive behaviour seeking to optimise performance according with the criteria along with evolution of training and expertise.

Adaptation, Self-organization and Error. It follows directly from this discussion that structuring the work processes through on-the-job training by an individual will be a self-organizing, evolutionary process, simply because an optimising search is the only way in which the large number of degrees of freedom in a complex situation can be resolved. The basic synchronization to the work requirements can based on procedures learned from an instructor or a more experienced colleague or it can be planned by the individual on occasion in a knowledge-based mode of reasoning by means of mental experiments. From here, the smoothness and speed characterizing high professional

skill together with a large repertoire of heuristic know-how rules will evolve through an adaptation process in which 'errors' are unavoidable side effects of the exploration of the boundaries of the envelope of acceptable performance. During this adaptation, performance will be optimised according to the individual's subjective process criteria within the boundary of his individual resources. This complex adaptation of performance to work requirements, eliminating the necessity of continuous choice will result in stereotype practices depending on the individual performance criteria of the agents. These criteria will be significantly influenced by the social norms and culture of the group and organization. Very likely, conflict will be found between global work goals and the effect of local adaptation according to subjective process criteria. Unfortunately, the perception of *process quality* can be immediate and unconditional while the effect on *product quality* of the choice of an actor can be considerably delayed, obscure and frequently conditional with respect to multiple other factors.

In a first encounter, when representation of work constraints is not present in the form of instructions from an experienced colleague or a teacher, and know-how from previous experiences is not ready, the constraints of the work have to be explored in a knowledge-based mode from explicit consideration of the actual goal and a functional understanding of the relational structure of the work content. For such initial exploration as well as for problem solving during unusual task conditions, opportunity for test of hypotheses and trial-and-error learning is important. It is typically expected that qualified personnel such as process operators will and can test their diagnostic hypotheses conceptually - by thought experiments - before actual operations if acts are likely to be irreversible and risky. This appears, however, to be an unrealistic assumption, since it may be tempting to test a hypothesis on the physical work environment itself in order to avoid the strain and unreliability related to unsupported reasoning in a complex causal net. For such a task, a designer is supplied with effective tools such as experimental set-ups, simulation programs and computational aids, whereas the operator has only his head and the plant itself. In the actual situation, no explicit stop rule exists to guide the termination of conceptual analysis and the start of action. This means that the definition of error, as seen from the situation of a decision maker, is very arbitrary. Acts which are quite rational and important during the search for information and test of hypothesis may appear to be unacceptable mistakes in hindsight, without access to the details of the situation.

Even if a human actor is 'synchronized' to the basic requirements of work by effective procedures, there will be amble opportunities for modification of such procedures. Development of expert know-how and rules-of-thumb depends on adaptation governed by subjective process criteria. Opportunities for experiments are necessary to find shortcuts and to identify convenient and reliable cues for action without analytical diagnosis. In other words, effective, professional performance depends on empirical correlation of cues to successful acts. Humans typically seek the way of least effort. Therefore, experts will not consult the complete set of defining attributes in a familiar situation. Instead it can be expected that no more information will be used than is necessary for *discrimination among the perceived alternatives for action* in the particular situation. This implies that the choice is 'under-specified' (Reason, 1986) outside this situation. When situations change, e.g., due to disturbances or faults in the system to be controlled, reliance on the usual cues which are no longer valid, will cause an error due to inappropriate "expectations." In this way, traps causing systematic mistakes can be designed into the system. Two types of errors are related to this kind of adaptation: The effect of the test of a hypothesis of salient cues and action which turn out negative, and

the effects of acts chosen from familiar and tested cues when a change in system conditions make the perceived set of alternatives unreliable.

Work according to instructions which take into consideration the possible presence of abnormal conditions that will make certain orders of actions unacceptable presents an example in which local adaptation is in conflict with delayed and conditional effect on the outcome. The be safe, the instruction may require a certain sequence of the necessary acts. If this prescribed order is in conflict with the actor's immediate process criteria, modification of the prescribed procedure is very likely and will have no adverse effect in the daily routine. (If, for instance, an actor has to move back and forth between several, distant locations because only that sequence is safe under certain infrequent, risky conditions, his process criterion may rapidly teach him to group actions at the same location together because this change in the procedure will not have any visible effect under normal circumstances).

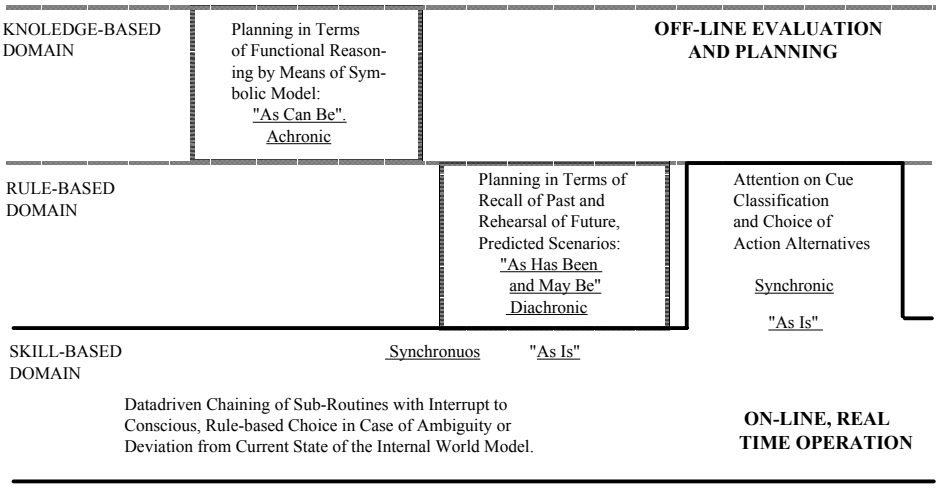

Figure 6 illustrates the complex interaction between the different levels of cognitive control. Tasks are frequently analyzed in terms of sequences of separate acts. In general, however, control of several acts takes place concurrent. Control of skilled sensorimotor activity is based on a continuous, dynamic interaction with the environment. Attention, on the other hand, is scanning across time and activities in order to analyze past performance, monitor current activity, and plan for foreseen future requirements. In this way, intuitive expectation is being prepared for oncoming demands and related cues and rules are rehearsed and modified to match predicted requirements. Symbolic reasoning is used to understand responses from the environment and to prepare rules for foreseen but unfamiliar situations. Attention may not always be focused on current activities, and different levels may simultaneously be involved in the control of different tasks, related to different time slots, in a time sharing or in a parallel processing mode.

Even within an established, effective sequence of *actions*, evolution of patterns of *movements* will take place according to subconscious perception of certain process qualities. In a manual skill, fine-tuning depends upon a continuous updating of automated patterns of movement to the temporal and spatial 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 tolerance limits, i.e. by the experience of errors or near-errors (speed-accuracy trade-off). Some errors, therefore, have a function in maintaining a skill at its proper level, and they cannot be considered a separable category of events in a causal chain because they are integral parts of a feed-back loop. Another effect of increasing skill is the evolution of increasingly long and complex patterns of movements which can run off without conscious

control. During such lengthy automated patterns attention is directed towards review of past experience or planning of future needs (see figure 6) and performance is sensitive to interference, i.e., capture from very familiar cues.

When delayed or conditional, global effects of behaviour are possible, error recovery by feedback correction and control of the local adaptation is not possible, and adaptation will be controlled by an evolutionary 'survival of the fittest' work process. In order to compete effectively with the effect of the local process criteria, the perception of fitness of such stored procedures must be maintained in another way (e.g., by artificial reinforcement or preferably, by rearranging the environment to include the global requirements in the local criteria). Otherwise, simple decay of memory of stored work rules (decay is, in effect, necessary for adaptation to changing requirements from a work environment) will necessarily require a repeated experience of the conflict, i.e., error, in order to maintain proper adaptation to characteristics of the environment.

The Structure of Cooperative work. So far, the discussion has been focused on the individual adaptation of work strategies to task requirement. In general, however, several people will be active in a work environment and also the adoption or allocation of the roles of the individuals evolve in a self-organizing mode according to local criteria and within the constraints of externally imposed allocation structures. Such constraints on the evolutionary role definition can have their origin in work requirements as well as in human resource limitations.

Role Allocation. Some constraints on work allocation originate in the work domain. Actions can, for instance, be required simultaneously in separate locations; or work can require competence which is depending on more than one profession. Such conditions will limit the extent to which allocation can be dynamically adapted to the preference of the involved individuals. In some cases, however, constraints are rather soft and will not be respected strictly during adaptation (e.g., the boundaries between activities which have been assigned members of different unions by labour market agreements). In other cases, constraints are effectively reinforced, for instance when performance is governed by strict quality control standards as is the case for manufacturing according to mill specs or in financial operations with strict legal control. In most cases, however, boundaries among the roles allocated the individual actors are continuously adjusted according to the requirements of the immediate work situation.

As it was the case for the choice among alternative work strategies, the dynamic shifting of boundaries between roles will be used to resolve resource demand conflicts and to match performance to individual preferences. The subjective criteria active in this adaptation will be very situation dependent and directly related to the particular work process, such as perception of differences in work load among colleagues, the amount of communication necessary among agents for coordination, subjective preferences for certain activities, etc. This adaptation of role allocation and coordination to work requirements during the normal conditions will endanger the functions during exceptional situations.

Coordination of cooperative work. For concerted work activity, the different processes and functions of work within the various levels of the means-ends space of a work domain will be allocated several individuals. Often, coordination will be allocated other individuals than those taking care of the functions to be coordinated. This is the case in all hierarchic organizations. In effect, boundaries are found between roles at different levels in the hierarchical control structure, as well as among roles within the levels.

The basic structure of the allocation depends on the functional requirements of the work content such as the topograhic location of work items, the work load related to certain functions, the timing required between functions in different places, and the time frames to consider in the coordination at the various levels. In other words, *technology shapes organization bottom-up* by posing strict constraints on allocation of functions to groups and individuals. In many domains, in particular in tightly coupled technical domains like manufacturing, process control, etc. strict control and timing requirements can be explicitly formulated from an analysis of the work requirements.

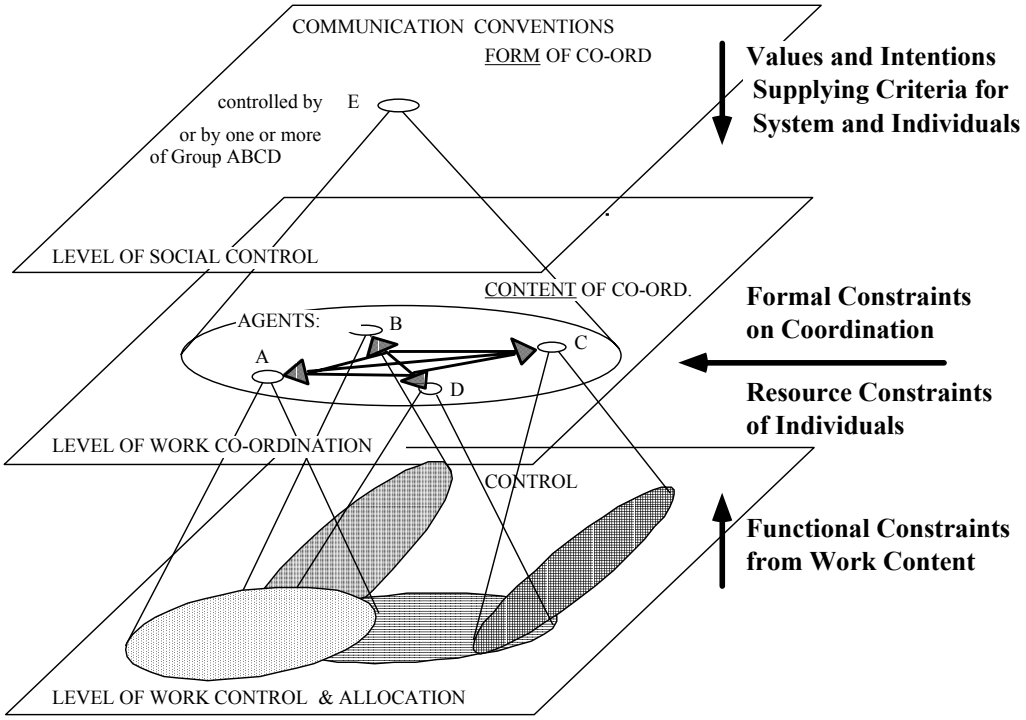

Figure 7 illustrates the coordination of cooperative work. At the level of work, a dynamically changing allocation to individuals is governed by criteria such as sharing load, minimizing communication, individual interest, etc. At the level of co-ordination, the *content* of communication necessary for concerted action is specified by the work content and the actual role allocation. In this way, the *work organization* is dynamically shaped bottom-up. Management practice and social values define rules of conduct, i.e., the *form* of the coordination and therefore, are shaping the *social organization* top-down. In addition, formal constraints such as laws, regulations, and union agreements add constraints on allocation and coordination 'side-in'. Within the boundaries defined in this way, there is plenty of room for adaptation guided by subjective criteria. (Adopted from Rasmussen, 1988c).

However, within the allocation and coordination constraints posed by the work content, there are many degrees of freedom to arrange the role allocation and to structure the way, coordination is brought about. Additional, *formal constraints* on allocation can originate in legal requirements (authorization, etc.), agreements (union boundaries), regulations (quality assurance standards) and rules of conduct (military).

System Reliability and Safety. The dynamic adaptation to the immediate work requirements both of the individual performance and of the allocation between individuals probably will be able to create a very high degree of reliability as long as the interaction

is transparent (i.e.; critical aspects are visible without excessive delay), and individual process criteria are not in conflict with or are not overriding critical product criteria.

System break-down and accidents are the reflections of loss of control of the work environment in some way or another. If the hypothesis is accepted that humans tend to close their degrees of freedom to get rid of choice and decision during normal work and that errors are a necessary part of this adaptation; the trick in design of reliable systems is to make sure that human actors maintain sufficient flexibility to cope with system aberrations, i.e., not to constrain them by an inadequate rule system. In addition, it appears to be essential that actors maintain 'contact' with hazards in a way that they will be familiar with the boundary to loss of control and will learn to recover. In 'safe' system in which the margins between normal operation and loss of control are made as wide as possible the odds are that the actors will not be able to sense the boundaries and, frequently, the boundaries will then be more abrupt and irreversible. When radar was introduced to increase safety at sea, the result was not increased safety but more efficient transportation under bad weather conditions. Will anti-blocking car brakes increase safety or give more efficient transport together with more abrupt and irreversible boundaries to loss of control? A basic design question is: How can boundaries of acceptable performance be established that will give feedback to a learning mode in a reversible way, i.e., absorb violations in a mode of graceful degradation of the opportunity for recovery.

Under certain conditions, however, self-organizing and adaptive features will necessarily lead to 'catastrophic' system behaviour unless certain organizational criteria are met. Adaptation will normally be governed by local criteria, related to an individual's perception of process qualities in order to resolve the perceived degrees of freedom in the immediate situation. Some critical product criteria (e.g., safety) are conditionally related to higher level combination or coincidence of effects of several activities, allocated different agents and probably, in different time slots. *The violation of such high level, conditional criteria cannot be monitored and detected at the local criterion level,* and monitoring by their ultimate criterion effect will be too late and unacceptable. Catastrophic effects of adaptation can only be avoided if local activities are tightly monitored with reference to a *prediction* of their role in the ultimate, conditional effect, i.e., *the boundaries at the local activities are necessarily defined by formal prescriptions, not active, functional conditions.*

This feature of adaptation to local work requirements probably constitutes the fallacy of the defence-in-depth design principle normally applied in high risk industries (Rasmussen, 1988d). In systems designed according to this principle, an accident is dependent on simultaneous violation of several lines of defence: an operational disturbance (technical fault or operator error) must coincide with a latent faulty maintenance condition in protective systems, with inadequacies in protective barriers, with inadequate control of the location of people close to the installation etc. The activities threatening the various conditions normally belong to different branches of the organization. The presence of potential of *a catastrophic combination of effects of local adaptation* to performance criteria can only be detected at a level in the organization with the proper overview. However, at this level of the control hierarchy (organization), the required understanding of conditionally dangerous relations cannot be maintained through longer periods because the required functional and technical knowledge is foreign to the normal management tasks at this level.

The conclusion of this discussion is that catastrophic system breakdown is a normal feature of systems which have self-organizing features and at the same time, depend on protection against rare combination of conditions which are individually affected by adaptation. Safety of such systems depend on the introduction of locally visible boundaries of acceptable adaptation and introduction of related control mechanisms. What does this mean in terms of organizational structures? What kind of top-down influence from 'management culture' and bottom-up technological constraints can be used to guide and limit adaptation? *How can we model and predict evolution of organizational structure and the influence on system safety?*

Conclusion

The conclusion of this discussion is that work in modern high-tech societies call for a reconsideration of the notion of human error and research should be focused on understanding human behaviour and social interaction in general *in cognitive terms* in complex, dynamic environments, not on fragments of behaviour called error. The approach has similarities to the risk homeostasis theories of traffic safety with the reservation that the controlling mechanisms are adaptation in a wider functional sense than control governed by criteria related to risk only. An important consequence of the general adaptivity of human behaviour and the relationship of errors to exploration of boundaries during adaptation is that the success of high-hazard/low-risk design principles like 'defence-in-depth' depends on precautions taken to operate the systems according to the basic assumptions behind the design philosophy. Such precautions in turn depend on the use of risk management based on predictive risk analysis (Rasmussen, 1988d).

The arguments presented should not be taken to be arguments against causal post accident analysis, predictive risk analysis, or the defence-in-depth design principle per se. On the contrary, these methods are indispensable tools in control of industrial safety. The point to consider carefully in the present rapid trend toward very large, tightly coupled systems is that we are faced with *new requirements for the use* of these tools, as it was argued above for the use of predictive risk analysis to operate 'defence-in-depth systems' according to their basic assumptions. Another example: A causal analysis of an accident can supply a very acceptable record of the course of events, it is a valuable set of data. Such a record, however, is *not a model of the system,* it does not represent the internal coupling among the local states in the system conditioning the flow of events. To predict the effect of improvements another analysis is necessary to identify the functional requirements and external pressures which are in fact controlling the adaptation leading to the new operational state of affairs. This kind of analysis is, at present, a research topic. A management error is not necessarily a human error which should be blamed on a manager, but depends on a deeper structural property which makes managers adapt to a particular type of behaviour. How can this deeper structure be identified?

The basic issue is that tightly coupled, large-scale systems require much closer coordination than found in the past in the development of the various tools used for design, operation, risk management, and regulatory measures.

References

FEINBERG, F. (1965): Action and Responsibility. In: M. Black (Ed.): Philosophy in America. Allen and Unwinn. Reprinted in: A.R. White (Ed.) (1968): The Philosophy of Action. Oxford Univ. Press

FITZGERALD, P.J. (1961): Voluntary and Involuntary Acts. In: A.C.Guest (Ed.): Oxford Essays in Jurisprudence, Clarendon Press. Reprinted in: A. R. White (Ed.) (1968): The Philosophy of Action. Oxford Univ. Press

KUHN, T. S. (1962): "The Structure of Scientific Revolution." University of Chicago Press, 1962.

LEPLAT, J. AND RAMUSSEN, J. (1984): Analysis of Human Errors in Industrial Incidents and Accidents for Improvement of Work Safety. Accident Analysis and Prevention Vol. 16, No. 2, pp. 77-88.

MACH, E. (1905). Knowledge and Error. English edition, 1976. Netherlands: Dordrecht, Reidel.

RASMUSSEN, J. (1988A): Human Error Mechanisms in Complex Work Environments. Reliability Engineering and System Safety, 22 (1988), pp. 155-69.

RASMUSSEN, J. (1988b): Coping Safely with Complex Systems. American Association for Advancement of Science, invited paper for Annual Meeting, Boston, Februar 1988;

RASMUSSEN, J. (1988c): Modelling Distributed Decision Making. Position Paper for International Workshop on New Technology and Distributed Decision Making, Bad Homburg, May 1988.

RASMUSSEN, J. (1988d): Safety Control and Risk Management: Topics for Cross-Disciplinary Research and Development. Invited Key Note Presentation in International Conference on Preventing Major Chemical and Related Accidents. In: IChemE Publication Series No. 110; Washington: Hemisphere Publishing Corporation, 1988.

RASMUSSEN, J. (1982): Human Errors. A Taxonomy for Describing Human Malfunction in Industrial Installations. Journal of Occupational Accidents, Vol. 4, Nos. 2-4 pp. 311-333

RASMUSSEN, J. AND VICENTE, K. (1987): Cognitive Control of Human Activities and Errors; Implications for Ecological Interface Design. Invited paper, International Conference on Event Perception and Action, Trieste, Italy, August 1987. To appear in: International Journal of Man-Machine Studies, in press.

REASON, J. (1986): Cognitive under-specification: Its varieties and consequences. In B. Baars (Ed.), The psychology of error: A window on the mind. New York: Plenum.

ROSCH, E. (1975): Human Categorization. In: N. Warren (Ed.): Advances in Cross-Cultural Psychology. New York: Halsted Press.

SWAIN, A.D. AND GUTTMANN, H.E. (1983): Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications. NUREG/CR 1278. Albuquerque, NM: Sandia Laboratories.