Coping Safely with Complex Systems

Rasmussen, Jens

Publication date:
1989

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

Risø-M-2769

RISØ

Risø-M-2769

Coping Safely with Complex Systems

Jens Rasmussen

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

Risø National Laboratory **Risø – M – 2769**

Title and author(s)

Coping Safely With Complex Systems

Jens Rasmussen

Date July 1989

Department or group

Information Technology

Groups own registration number(s)

R-3-89

Project/contract no.

Pages 13 Tables Illustrations References 20

ISBN 87-550-1497-6

Abstract (Max. 2000 char.)

The general dependence on large scale systems together with rapidly changing technology require predictive models of the performance of complex systems in order to be able to judge in advance the functionality and safety of new system concepts. Complex systems including human actors, however, cannot be modelled by quantitative, deterministic models and causal models in terms of objects and events have typically been adopted.

The paper presents a discussion of several basic difficulties with this approach. Definition of human error during supervisory tasks is becoming increasingly difficult, and post-hoc identification of causes of an accident depends on a pragmatic stop-rule for the termination of the analysis. Empirical verification of the design of a complex system, likewise, raises the question of stop-rules for adjusting the experimental conditions.

In addition, there is a need for the development of predictive models of human performance in complex systems. For intellectual, creative tasks, they cannot be in terms of procedural task descriptions but have to be based on higher level 'first principles' which take into account goal oriented, adaptive human characteristics. This approach is discussed for models of individual human actors as well as for models of organizations managing large scale systems.

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-2769

COPING SAFELY WITH COMPLEX SYSTEMS

Jens Rasmussen

ABSTRACT

The general dependence on large scale systems together with rapidly changing technology require predictive models of the performance of complex systems in order to be able to judge in advance the functionality and safety of new system concepts. Complex systems including human actors, however, cannot be modelled by quantitative, deterministic models and causal models in terms of objects and events have typically been adopted.

The paper presents a discussion of several basic difficulties with this approach. Definition of human error during supervisory tasks is becoming increasingly difficult, and post-hoc identification of causes of an accident depends on a pragmatic stop-rule for the termination of the analysis. Empirical verification of the design of a complex system, likewise, raises the question of stop-rules for adjusting the experimental conditions.

In addition, there is a need for the development of predictive models of human performance in complex systems. For intellectual, creative tasks, they cannot be in terms of procedural task descriptions but have to be based on higher level 'first principles' which take into account goal oriented, adaptive human characteristics. This approach is discussed for models of individual human actors as well as for models of organizations managing large scale systems.

July 1989

Risø National Laboratory, DK 4000 Roskilde Denmark

Invited paper for American Association for the Advancement of Science,
AAAS annual meeting, Boston MA, 11-15 February, 1988

ISBN 87-550-1497-6
ISSN 0418-6435

Grafisk Service, Risø 1989

CONTENTS

INTRODUCTION

Science and engineering depend on a representation of the laws of nature in control of the behaviour of physical systems. This representation can take different forms. The classical representation from Aristotle and onward was formed in terms of causal connections between events. The breakthrough of modern science was due to Galilee and Newton who replaced observations of events by measurements of variables and causal laws by mathematical relations among variables.

The quantitative, mathematical representation of the physical sciences and engineering has been so successful that the qualitative concept of causality has been discredited by scientists. In his classical essay on the notion of cause, Russell (1912) finds the concept of causality to be so diffuse that it should be banished from science. Russell's conclusion, that causal explanations should be replaced by relational, mathematical representations appears, however, to be mistaken. The two methods of representation are complementary approaches and they are both necessary for engineering analysis.

The quantitative, relational representation of physical sciences is not applicable for analysis of the courses of events when the structure of technical systems breaks down, e.g., during accidents. Likewise, this representation is not suited to describe the interaction of human decision makers and technical systems. The consequence is that a representation in terms of causal flow of events has become an important tool for accident analysis and for modelling decision making in the control of technical systems.

Causal explanations describe objects which interact in chains of events. Neither the objects nor the events can, however, be defined objectively. Their identification depends on a frame of reference which is taken for granted and causal explanations are only suited for communicating among individuals having similar experience who share more or less intuitively the underlying definitions. In a period of rapidly changing technology and the involvement of laymen in arguments about the impacts of large scale technical installation, the ambiguity caused by the very nature of causal explanations is an important problem.

RELATIONAL AND CAUSAL REPRESENTATIONS

The causal and the relational representations are supplementary. They are based on fundamentally different methods of generalization and will serve different purposes for scientific and engineering analysis.

A mathematical, relational representation of physical phenomena is based on mathematical equations relating physical, measurable variables. The generalization depends on a selection of relationships which are 'practically isolated' (Russell, op. cit.). This is possible when they are isolated by nature (e.g., being found in the planetary system) or because a system is designed so as to isolate the relationship of interest (e.g., in scientific experiment or a machine supporting a physical procession a controlled way). In this representation, material objects are only implicitly present in the parameter sets of the mathematical equations. The representation is particularly well suited for analysis of the optimal conditions and theoretical limits of physical processes in a technical system which, by its very design, carefully separates physical processes from the complexity of the outside world.

A causal representation is expressed in terms of regular causal connections of events. In his essay, Russell discusses the ambiguity of the terms used to define causality: the necessary connection of events in time sequences. The concept of an 'event,' for instance, is elusive: the more you strive to make the definition of an event accurate, the

less is the probability that it is ever repeated. In this way, the regularity of causal connections disappears when attempts are made to define the concepts objectively. The weakness of Russell's request to have causal concepts defined objectively is its root in the quantitative, mathematical representation: In order to qualify the argument that a stone thrown against a pane of glass breaks it, you have to specify the weight and velocity of the stone. This argument is basically wrong. Events and causal connections cannot be defined by lists of objective attributes. An attempt to qualify a causal statement objectively by events in conjunction with the conditions which are jointly sufficient and individually necessary for a given effect to occur is, as Russell observed, without end. Completeness removes regularity. The solution is, however, neither to give up causal explanations, nor to seek objective definitions. Regularity in terms of causal relations is found between kinds of events, not between particular, individually defined events.

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, prototypes (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 and it has been met in many attempts to design computer programs for picture analysis. The 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.

A classical example is "the short-circuit caused the fire in the house" (Mackie, 1965). This statement in fact only interrelates the two prototypes: the kind of short-circuit that can cause afire in that kind of house. The explanation that the short-circuit caused a fire may be immediately accepted by an audience from a region where open wiring and wooden houses are commonplace, but not in a region where brick houses are the more usual kind. If not accepted, a search for more information is necessary. Short-circuits normally blow fuses, therefore further analysis of the conditions present in the electric circuit is necessary, together with more information on the path of the fire from the wiring to the house. A path of unusually inflammable material was probably present. In addition, an explanation of the short-circuit - its cause - may be needed.

The explanation depends on a decomposition and search for unusual conditions and events. The normal and usual conditions will betaken for granted, i.e., implicit in the intuitive frame of reference. Therefore, in causal explanations, the level of decomposition needed to make it understood and accepted depends entirely on the intuitive background of the intended audience. If a causal statement is not accepted, formal logical analysis and deduction will not help, it will be easy to give counter-examples which can not easily be falsified. Instead, further search and decomposition are necessary until a level is found where the prototypes and relations match intuition.

In the same way as it is impossible to define the meaning of words by linguistic analysis of one separate sentence, it is impossible by analysis to define the elements in a causal statement separated from its verbal context in the total description. In effect, causal explanation are only suited for communication among individuals who share prototypical definitions of objects and events because they have similar experience and, therefore, common 'tacit knowledge' (Polanyi, 1967). The great effort spent to formalize

causality and to cope logically with counterfactual statements (Sosa, 1975) is probably misdirected. Instead efforts should be focused on recent linguistic developments such as relevance theory (Sperber and Wilson, 1986) and situational semantics (Barwise and Perry, 1983) and on psychological theories of conceptualization and categorisation (for a review see Murphy and Medin, 1985).

The causal description is an analog representation including physical objects as separate elements. The generalization implies categorisation and identification of prototypical objects and events. The great value of causal reasoning is its immediate relationship to the material world, i.e., to physical objects and their configuration. The representation is therefore very easy to update in correspondence with changes in the real world. This is not the case in the relational representation in which a complex set of parameters and variables must be changed in order to incorporate physical changes.

In this way, the qualitative, causal reasoning is useful to guide reasoning during design or in a choice of physical systems for some purpose. On the other hand, the mathematical reasoning, related to formal analysis of relations between variables, is particularly useful to optimize a design and find its theoretical limits. The complementary nature is similar to that found between the use of intuitive judgement and formal proof by mathematicians (see Hadamar 1945).

PROBLEMS IN ANALYSIS OF MODERN TECHNOLOGY

In the more traditional use of technology, the problems related tote ambiguities of causal representations were rather innocent. Quantitative engineering analysis was applied for the design of machinery. A classical steam locomotive could be considered by the designers to be a well-defined micro-world in which the relationships of the thermodynamic laws were undisturbed by external factors. The description of the interaction of the locomotive with the environment in the course of an accident was the concern of others who could then view it as one integrated object with certain characteristics. Its behaviour in the environment could be described with no reference to its internal physical functioning.

This separation cannot be maintained for the engineering analysis of large, centralized systems. Large systems present potential for great losses and damage to people and environment in case of internal malfunction. Systems such as large chemical process plants cannot be considered to have 'practically isolated' internal functions, well contained by system boundaries and, therefore, adequately described by classical engineering analysis. Accidents happen when system boundaries break down. In this case, the preconditions for formal, mathematical analyses of system function also break down and the formal methods are replaced by different methods for analysis of accidents based on causal representations.

Such causal representations are important for two purposes. One is the analysis of accidents and incidents in order to gain experience and to collect data for the improvement of the safety of future designs. Another purpose is risk analysis, i.e., predictive analysis of hypothetical courses of events following technical faults and human errors.

PROBLEMS IN CAUSAL ANALYSIS OF ACCIDENTS.

Analyses of accidents in, e.g., chemical process plants, are made 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. A direct reference to elements in the physical world makes causal analysis a very effective technique for identifying and representing accidental

conditions. It is, however, important to consider the implicit frame of reference of a causal analysis.

In the analysis of accidents, decomposition of the dynamic flow of changes will normally terminate when a sequence is found with events which match the prototypes familiar to the analyst. The resulting explanation will take for granted his frame of reference and only what he find is unusual generally 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 may be set up by abnormal conditions which are latent effects of earlier events or acts. 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 inquest ion in the present context is: how does the degree of decomposition of the causal explanation and the selection of which side-branches to include 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.

The perception of stop-rules is very effective 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 s□□top-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) when a familiar, abnormal event is found to be a reasonable explanation; or (c) if 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 may very likely reach different conclusions.

To summarize: identification of accident causes is controlled by pragmatic, subjective stop-rules which to a large extent also 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.' 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 trained better, a further search would probably be made. In such a case, a design or a manufacturing error can be found.

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 oft he very nature of the causal explanation. Human error is, particularly at present, familiar to analysts: to err is human, and highly skilled people will frequently depart from normative procedures. The problem of defining human error will be discussed in a separate 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 if 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 psychological factors also should be taken 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 who will be more 'in power of control' than an operator in the dynamic flow of events. These 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 or because they are to be found in a conditioning side branch of the causal tree, not directly involved in the dynamic flow.

In conclusion, present technological developments require 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. The present concept of 'power of control' should be reconsidered from a psychological point of view, as should the ambiguity of stop-rules in causal analysis.

Analysis for System Improvements.

Analysis for therapeutic purpose, i.e., in order to identify events or conditions which can be target 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 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 where it would be psychologically feasible to ask for a change in behaviour when reports from accidents focus only on the flow of unusual events?

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.

HUMAN ERROR, A STABLE CATEGORY?

The causal representation is generally used for the description of human activities. Analysis of manual tasks results in time sequences of observable acts. For mental tasks, a well-known representation is the Newell-Simon production system with rules linking cues with actions. Actually, however, human activity frequently has the character of a smooth, dynamic behaviour. Task analysis in this case require a somewhat artificial decomposition into acts together with identification of their cues. In addition, to make a description complete, certain kinds of acts re-classified as errors.

There will be no great ambiguity in identifying human errors from routine tasks. They can normally be decomposed into separate, observable acts which are, to a large extent, controlled in sequence by the physical work content. In addition, such tasks are repetitive, and the actors have reached a stable level of skill. Therefore, "normal performance" is known and can be used as a reference for identification of errors.
Automation of industrial systems has replaced many repetitive tasks by supervision and trouble-shooting. Such tasks depend on cognitive processes for diagnosis and contingency planning. Performance is then much less constrained by the external task conditions and can based on several different strategies. Furthermore, a stable level of training can no longer be assumed. In this situation, errors are not members of a stable prototypical category defined with reference to normal behaviour. They will be identified by their effects in analyses aimed at explaining and/or to finding the cause of unsatisfactory system performance.

When performance can no longer be judged with reference to a stable, normal performance, the definition of "human error" becomes dubious because of the ambiguity of the stop rule. It is the fate of the people involved in accidents that everybody in hindsight (usually lacking detailed information about the context of the actions in question) can imagine a cure for human errors in terms of more care, better training or instruction, or direct punishment. Paradoxically, human errors seem to be the source of blame in two contradictory circumstances. On one hand, human errors are found when normal human variability occasionally brings task performance outside acceptable limits. On the other, human errors are found when human variability or adaptability proves insufficient to cope with variations in task content; i.e., if it is found, on hindsight, that a "reasonable" human ought to be able to cope with disturbances. Basically, human error is related intimately to a mismatch between two varying partners, a human and the work requirements. Errors can not be defined in terms of stable categories.

Human error and learning.

Human errors appear to be very closely related to the learning process during adaptation to the requirements of a task.

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 optimization 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.

Also development of expert know-how and rules-of-thumb depends on a basic variability of behaviour. 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, it can be expected that no will be used than are necessary for discrimination among the perceived alternatives for action in the particular situation. This implies that the choice is 'under-specified' (Reason, 1987)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.

An important issue is that error mechanisms cannot be separated from adaptive processes which are very useful. Reason (1987)stresses a similar view: He finds that examination of a wide variety of error forms suggests that they reflect basic error tendencies. "These tendencies, it is argued, constitute the root of most, if not all, of the systematic varieties of human error. Each of these error tendencies is necessary for normal psychological functioning. It is from this necessity that they derive their great power to induce systematic error." The conclusion of this discussion is that the cue-action relationship cannot be considered a stable causal connection, but depends on subjective features in the experience of an individual actor.

For problem solving during unusual task conditions, test of a hypothesis becomes an important need. It is typically expected that operators check their diagnostic hypotheses conceptually - by thought experiments - before operations on the plant. 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 precise 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, in hindsight without access to the details of the situation, appear to be unacceptable mistakes.

In conclusion, the very nature of human error depends equally on normally very efficient cognitive mechanisms and particular features in the work context created by system designers and managers. The rapid technological change together with the ambiguity of causal analysis, therefore, make it very important to reconsider the traditional view of error and responsibility.

DECISION MAKING IN ADVANCED SYSTEMS.

The relationship between human error and adaptation to the requirements of the work increases the responsibility of designers of large-scale systems and the need for a re-formulation of causal representations.

The large potential for loss and damage from accidents in advanced systems has resulted in a particular design practice. It is not acceptable that single component failures or human errors can release a chain of events leading to accidents and losses and therefore a design philosophy of 'defence-in-depth' has evolved. The result is that

systems have numerous lines of defence such as protective functions, barriers against fault propagation, etc., which can serve to terminate accidental chains of events before serious loss and damage can occur. In addition, stand-by equipment is installed and is supposed to take over when operating systems fail. A disturbance, e.g., a fault or human error, can then only evolve into an accident when it coincides other faults that make the safety measures inactive.

In such a system, many errors and faults made by the operating staff and maintenance personnel do not directly reveal themselves through an observable functional response from the system. Adaptation to task requirements which are very reasonable in the immediate context can therefore violate safety features without visible effect for the actor. It is easy, after the fact, to identify unacceptable violations of a design-in-depth design concept. Seen separately, however, the violations can be reasonable and, in fact, necessary for the flow of work. What should be controlled is the possible coincidence, not the individual act. Such a control depends on an overview of the causal structure of the risk analysis, not only of the actual work situation. Whether they are in fact unacceptable errors or violations depends upon a possible coincidence with other conditions which are invisible to the actor. Compare the classic dictum of Mach (1905): "Error and knowledge flow from the same source, only success will tell the difference." Acceptability of an act simply depends on the emphasis put on different conflicting criteria of judgement. An emphasis which will be very different in the judgement of the actor faced with the actual work requirements and for whom the possible effect on preconditions of safety can be obscure and, after the fact, in the judgement of the accident investigation committee.

In this kind of system, it is mandatory that the preconditions oft he causal analysis of designers are made explicit and that they are understood and accepted by the operating staff and management who, unfortunately, have different 'tacit knowledge' than designers. The dependence of causal analysis on shared prototypes and frame of reference and the related problem of communication between different professions is now becoming an important problem for practical risk management in hazardous industries (Rasmussen,1987).

PROBLEMS IN CAUSAL RISK ANALYSIS

In reports of industrial risk analyses, reference is typically made to the fault-tree and event-tree methods. It is usually not realized that fault trees and event trees do not represent methods of analysis, but are merely records of their results. To qualify the results, the context from which the causal trees are extracted and the way in which they are identified should be made explicit.

This is generally possible for the propagation of the effects of component faults in the plant itself. Technical components are well defined object prototypes with widely accepted failure modes, and the course of events is largely dictated by the plant anatomy. A rather well-defined completeness can, therefore, be obtained by hazard identification methods following all causal paths found in the pipe-and-instrumentation diagram of a plant.

It is much more difficult to take the human influence on accidents into account. Objects, then, include mental concepts, events include decisions and actions, and the course of events will depend on human communication and mobility. In this case, there is no well defined and stable physical structure to guide the predictive modelling of the causal chains which could be significant contributors to risk. It is therefore very difficult to state explicitly the completeness of a causal analysis including human activities. The high risk scenarios to consider during design will have to include

complex situations caused by several coinciding events. The identification of possible causal networks in advance requires the equivalent of a design algorithm:Identify potential sources of accidents in terms of accumulations of energy together with the possible victims and targets; then 'design' the accident by identifying the physical functions and human actions which are necessary and sufficient for its release;and, finally, find those activities which, in case of errors and faults and proper timing, can be changed so as to match the identified accident patterns.

Unfortunately, such causal prediction is as open-ended as is any other design task. There is, at present, no way to state its completeness explicitly. Development of basic principles from which causal patterns can be generated in a systematic way is very important. Without these, the safety of large systems will depend on a continuous and ad-hoc back fitting to include new precautions based on the latest accident.

EVALUATION OF THEORIES AND SYSTEMS.

The internal consistency of a quantitative, relational model can be tested logically and mathematically, and systems designed from such theories can be validated by means of controlled experiments. Implementation of these models in the real world is considered to be an 'application' and, therefore, irrelevant to science. Such rigid distinctions are not appropriate when considering theories relating to complex environments and based on causal representations obtained from the generalization of observations.

Artificial intelligence research offers effective tools for simulation of complex systems by object oriented languages. This has created a new 'cognitive science' in which theories are only acceptable, if they can be tested for consistency by computer simulation. Until now, this has only been successful for very well-formed micro-worlds like games, cryptograms, and theorem proving. The basic reason for this appears to be that objects, events, and causal connections in such restricted worlds can be objectively defined in the classical way by lists of attributes.

This is not possible for simulation of complex systems. Objects, events and causal relations in models then represent classes, not instances. It is, of course, possible in simulations to replace the classes by particular members, but then the entire exercise will be an ad-hoc demonstration of selected examples. There will be no formal stop rule to terminate the additions of new relations or objects in order to match simulated performance to observed real life performance. An empirical testing of the internal consistency of a theory in causal terms by simulation of scenarios is not a satisfactory solution. As Davidson (1967) suggested, causal descriptions and causal laws should be kept apart from each other.

Simulation of causal chains of events in entirely technical systems is eased by their well structured and relatively stable anatomy. Simulation can be planned from invariant relationships at a higher level of abstraction, typically derived from mass and energy conservation laws. Classes of events related to state changes of physical components can be mapped onto model parameters and the completeness of representation of a set described by a prototype can be judged.

This is not the case for the activity of people with free will, mobility and subjective goals. During actual behaviour, members of a causal chain will be selected from prototypical classes and adjusted to match particular conditions. In simulation based directly on explicitly formulated causal relations, e.g., production rules, this is not the case. The set included in the data base of a simulation, therefore, has to include all possible variants of a class.

This is unrealistic; consider for instance the variety of actions necessary to represent the class of 'human errors'. Simulation based on 'first principles' corresponding to physical conservation laws is therefore necessary to avoid ad-hoc selection and model fitting. For human error, this implies that particular errors should be generated from higher level behavioural laws.

A first approximation is to categorize error according to generating mechanisms: effects of adaptation guided by the law of least effort; interference between cognitive control structures;lack of mental processing resources, etc. When such principles underlying observed scenarios can be formulated, an 'object-oriented' simulation can 'generate' entire families of scenarios from higher level relationships.

Similarly, representing human knowledge in terms of explicit 'production rules' appears to involve a category mistake, taking the particular for the kind. Taken as models of human expertise, the old GPS-system was more valid than the recent expert-system approach. Advice from an expert system will probably only be understood and accepted by a person who has the same frame of reference as the expert supplying the rules, and therefore is able to regenerate the prototypical classes (i.e., in practice; only the expert himself?).

In order to avoid ad-hoc demonstration in simulation of human decision making, independent representations are necessary of the work domain, the task requirements and of the human cognitive mechanisms. This, in fact, is a request for the revival of the Brunswikian ecological psychology (Brehmer, 1984).

Another example is simulation of organizational behaviour. Decision making is typically modelled in terms of a separate sequence of mental acts of one person. Observations are made, an analysis supports a diagnosis and, finally, acts are planned. In reality, however, actual decision making is a continuous process to control the state of affairs in a dynamic work domain in co-operation with other decision makers.

Models of organizations are, however, normally based on normative, effective decision strategies, and the co-operation on the role allocated in formal organizations. Again, simulation based on such representations will be ad-hoc demonstrations. Instead, simulation should be based on principles that will dynamically generate the decision scenarios. One approach can be from a control theoretic point of view in terms of a self organizing, distributed control system in which the actors are allocated control of a defined sub-space of a loosely coupled work domain defined by the available means-ends relations. The actual organization of actors can then evolve from a specification of the conventions and constraints defined for their co-operation and communication.

CONCLUSION

During the past period of slowly evolving technology, the ambiguity of the causal representation of common-sense reasoning was a rather innocent academic problem of philosophy and human sciences. In a rapidly developing technology, change inhibits the maintenance of common experience and intuition of groups involved in the development, use, and assessment of technology. At the same time, design of large complex systems makes the use of causal analyses mandatory. Basic research and application can no longer be separated.

ACKNOWLEDGMENT

The problems discussed in the present paper emerged during work on improvement of work safety supported also by the EEC programme of Ergonomics for the E.C.S.C In-

dustries, Luxemburg. The co-operation and discussions during this project with Jacques Leplat, Paris are gratefully acknowledged.

REFERENCES

BARWISE, J. AND PERRY, J. (1983): Situations and Attitudes. Cambridge, Mass: MIT Press.

BREHMER, B. (1984). Brunswikian psychology for the 1990s, in: K. M.J. Lagerspetz and P. Niemi, (Eds.) Psychology in the 1990s, New York: Elsevier Science Publishing Co.

DAVIDSON, D. (1967): Causal Relations. Journal of Philosophy, Vol.64, pp. 691-703. Reprinted in: E. Sosa (Ed.) (1975): Causation and Conditionals, Oxford University Press.

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

HADAMAR, J. L. (1945): "The Psychology of Invention in the Mathematical Field." Princeton University Press.

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

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

MACKIE, J. L. (1975): "Causes and Conditions." American Philosophical Quarterly, Vol. 2.4 pp. 245-255 & 261-264 Reprinted in: E. Sosa (Ed.) (1975): Causation and Conditionals, Oxford University Press.

MURPHY, G. L. AND MEDIN, D. L. (1985): The Role of Theories in Conceptual Coherence. Psychological Review, Vol. 19, No. 3, pp. 289-316

NEWELL, A., SHAW, J. C. AND SIMON, H. A. (1960): "Report on a General Problem-Solving Program for a Computer." Information Processing: Proceedings of the International Conference on Information Processing, pp. 256-264. Paris: UNESCO.

NEWELL, A. AND SIMON, H. A. (1972): "Human Problem Solving." Prentice Hall, New Jersey.

POLANYI, M. (1966): "The Tacit Dimension." New York: Doubleday.

RASMUSSEN, J.: Approaches to the Control of the Effects of Human Error on Chemical Plant Safety. International Symposium on Preventing Major Chemical Accidents. February 1987, American Institute of Chemical Engineers.

REASON, J. (1986, Personal Communication): Cognitive Under-Specification: Its Varieties and Consequences. To be published in B. Baars (Ed), The Psychology of Error: A Window on the Mind. New York: Plenum

REASON, J. (1988, Personal Communication): Human Error. New York: Cambridge University Press. In Preparation.

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

RUSSELL, B. (1913): "On the Notion of Cause". Proc. Aristotelean Society, Vol. 13, pp. 1-25.

SOSA, E. (1975) (Ed.): Causation and Conditionals, Oxford University Press.

SPERBER, D AND WILSON, D. (1986): Relevance. Oxford: Blackwell