Decisions in complex systems; finding a balance between competing forces

Björn Wahlström
Technical Research Centre of Finland (VTT)
POB 1000, FIN-02044 VTT (Espoo)
bjorn.wahlstrom@vtt.fi

Abstract. Disastrous accidents have shown that small deficiencies and shortcomings can lead to large losses in terms of human life, environmental destruction and economic values. Risk research has identified human and organisational factors as the most important set of underlying causes to such accidents. This has made it increasingly evident that decisions taken by managers at different levels in the organisations have a very large influence on the initiation and outcome of such events. The paper discusses models of decision making and complex systems in a search for some kind of "super-rational" model of decision making, which would provide a better support in decision making situations that occur in the control of high-risk technologies. The concept of interconnected systems of intelligent autonomous agents may be suggested as a candidate for such a model. The drawback is however that such a model can be used only in a qualitative and metaphoric sense for understanding safety-oriented organisations and not for an accurate prediction of their behaviour.

1 INTRODUCTION

Disastrous accidents in different fields such as Three Mile Island, Chernobyl, Piper Alfa, Challenger, Columbia, etc., have shown that small deficiencies and shortcomings can lead to large losses in terms of human life, environmental destruction and economic values. Risk research has identified human and organisational factors as the most important set of underlying causes to such accidents. Already a simple analysis shows that choices and decisions made by managers at different levels in the organisations have a large influence on the outcome of such events both in the good and in the bad.

The overreaching question is thus how these choices and decisions can be supported by methods and tools to avoid the triggering of unwanted events and to introduce better control and mitigation of such sequences of events when a trigger occurs. Research in human decision making has suggested various models by which the quality of decisions can be improved, but the models are difficult to apply in practice. One possible solution to the modelling problem could be some kind of "super-rational" model of decision making, which would provide a general frame of understanding of problems that emerge in the control of complex technological systems.

The difficulty in finding such a general model is that it has to integrate many different models, which form the base of our present understanding of problems and available remedies. It has to be multi-disciplinary and it should expose decision makers to the need for balancing several apparently conflicting views to create strategies that can be communicated, accepted and implemented in the organisations in consideration. The discussion below starts with a short account for the most important concepts within decision theory and moves over to our present understanding of human decision making. A short description of a few important concepts of complex systems serves as an introduction to a more elaborate

Presented at the 26th NeTWork Workshop on "Resolving Multiple Criteria in Decision-Making Involving Risk of Accidental Loss", Schloss Steinhöfel, Germany, 27 – 29 September 2007.

discussion of high risk technologies. Possibilities for building better decision models are considered before the conclusions of the paper.

2 A BASE IN DECISION THEORY

Decision theory has for the last fifty years developed to be an own area of science, which is based on mathematics and behavioural sciences. Decision theory has found its applications widely within business, industrial and military applications.

2.1 Optimisation

The foundations of decision theory lie in optimisation, where optimal solutions to certain mathematical problems are found through a consideration of necessary and sufficient conditions for optimality. A decision problem can generally be seen as consisting of the following four components:

- a set of possible decisions,
- a system, which associates an outcome to each possible decision,
- a utility functions, which calculates the utility connected to each pair of decision and corresponding outcome,
- additional restrictions connected for example to decisions and outcomes that are either enforced or not allowed.

In this simple formulation the optimisation problem contains an implicit difficulty in how outcomes should be connected to utility. In many practical cases outcomes are directly proportional to the utility, but in other cases non-linear functions have to be used. The solution to the St. Petersburg paradox suggested by David Bernoulli is for example using a logarithmic function to tie utility to money. Another difficulty is that utility functions are seldom stable over time and different people may have very different utility functions.

Most decision problems involve systems that evolve dynamically in time and then the utility function has therefore to reflect utilities that are collected over time. A simple example is the need to invest in near time for utilities that are accrued sometime in the future. The utility function has thus to integrate costs and benefits over time, but the difficulty is to define this time frame and applicable discounting principles.

2.2 Uncertainty

The need for handling of uncertainty arises in most decision problems and it is usually handled by introducing a probability function over the set of outcomes. This extension of the simple optimisation problem gives the so called expected utility theory (EUT). Criticism to EUT has been expressed among others by Maurice Allais in pointing out the human preference for certainty. Humans also typically show a non-symmetric attitude to gains and losses, which was addressed by Amos Tversky and Daniel Kahneman in their prospect theory. Other similar cases, where human decisions do not fulfil the assumptions of the EUT axioms have been observed.

Including uncertainty in a decision model implies that probability function either is known or can be estimated. It is also important to note that probability in the usual sense only is one pf several possible models of uncertainty. Probability can for example be interpreted in different ways and there are other similar mathematically consistent models, which for example instead of probability talk about possibility. Fuzzy sets, which were introduced by Lotfi Zadeh, provide still another method that can be used to model uncertainty in logic statements.

2.3 Multiple attributes

In many real decision situations there is not only one, but instead multiple utility functions, which should be optimised at the same time. The multi-attribute utility theory (MAUT) can be seen as an extension of the EUT, but it is difficult to solve in the general case. A straightforward approach to MAUT is to construct some function that reduces the attributes to a single attribute. The most common function for this purpose is to form the utility to be optimised as a weighted sum of the attributes.

The problem with this straightforward approach is to assign the weighting coefficient for very disparate attributes. In spite of this difficulty the MAUT framework is often used in a process were the decision maker selects a set of criteria according to which the relevance of the attributes are assigned. If this process is used iteratively it can usually help in reaching decisions, which give a satisfactory balance between the attributes.

2.4 Two or more actors

The possibility of multiple players is given a formal concretisation in game theory. Zero-sum games between two persons were solved by John von Neumann and Oskar Morgenstern in the 1940's, but non-zero-sum games and games between more than two actors that may form coalitions have not yet got an exhaustive theoretical treatment. Also the possibility of a repeated play of non-zero-sum games has shown to contain issues that make the theoretical treatment of the games more complicated.

Negotiations, where two actors in a sequence of asks and bids, try to reach a mutually accepted compromise have shown to contain its own theoretical difficulties. A successful result of the negotiation process is that it can reach a point that is Pareto optimal, but this is not always true for real negotiations. To reach better models for negotiation processes it would for example be necessary to reflect how much the negotiators trust each other. The involvement of a third neutral actor in a mediator's role may in some cases ease the negotiation process.

2.5 Difficulties in applying decision theory to real decisions

Before a certain decision model can be used for real decisions, the actual situation has to be described formally in a suitable frame. This means that there is a trade-off between the resources that are spent on finding a good decision and the expected gains in improving the quality of the decision. The time allowed to make the decision is also restricted in real life situations.

In practice this means that real life decisions always have to be made with the additional utilities of time and resources in mind. If a decision is critical it would therefore be wise to spend more time and resources on it than on less critical decisions. Similarly if a decision is very urgent the best strategy may be to select the first path of action that comes into mind. The framing of a decision problem can therefore be seen as a decision problem in itself.

3 HUMAN DECISION MAKING

Decision theory has been developed from descriptive approaches towards normative and prescriptive theories in a search for models of rationality. Failures of human decision making to fulfil the predictions of these theories have helped in identifying too large simplifications of the underlying models. The behavioural sciences have had an important part in this endeavour.

3.1 Individual decision making

From an evolutionary point of view the most important decision of all our ancestors have been to flee or fight when exposed to an unexpected situation. This adaptation has certainly set its own traces in our ability to make split second decisions in critical situations. However, one may also assume that our ancestors have approached decision making problems of another kind, where long contemplations have generated sudden insights of how some artefact could provide a solution to some practical problem.

In considering human decision making one may separate between simple choice and decisions. Everyday actions typically evolve almost unconsciously along paths that are selected without identifiable decisions. One could in this case speak about simple stimulus – response type models that are controlling human behaviour. This smooth behaviour is the result of an evolutionary adaptation of ancient man to complex environments in which survival was the major driving force. It seems to be an adaptation that provides an efficient use of attention and other resources to the responses that are needed in an ever changing environment.

Decisions can be sub-divided into smaller parts, where the need for a decision is detected, information is collected, possible actions are evaluated, their outcomes are assessed, actions are planned and executed and finally the outcomes are followed. These parts can actually be seen as mini-decisions in their own in which the decision maker may spend more or less mental efforts. This sequential model was developed into a separation between knowledge, rule and skill based behaviour, where the degree of automation of a certain task gives a prediction of the attention and resources needed and typical errors that are made [1].

3.2 Decision making in groups

The co-operation in small groups has apparently been one of the key factors in the evolution of ancient man. Game theory has however shown that co-operation mostly is interwoven with a component of competition. One can only speculate how these two components interacted to form the modern man, but co-operation relies on co-ordination, which calls for the appointment of a leader of the small group.

Group behaviour has been studied in social psychology and interesting dynamics evolves when members of a group interact. The behaviour of a group is determined both by cognitive processes of the individuals and by social processes within the group. Most importantly these processes interact and create social norms that influence group behaviour. Emotions and empathy also seem to be important factors in predicting behaviour of a small group.

A typical behaviour of group is that their members take up different roles. This division of labour has the benefit of making the group more efficient, but it also brings a source of conflict. Are all roles equally valuable for the group or is it fair that some group member gets a higher reward for his or her work? The rules for group formation and dissolution may also influence group behaviour, especially if members have the possibility to resign from the group at will.

Organisations are governed by mechanisms of social behaviour. Organisations typically have some kind of hierarchical structure, which means that selection of leaders on different levels will be an important factor in the co-ordination of organisational efforts. How are the opinion of these leaders valued and how are their desires responded to? Such issues become the organisational culture or the fabric, which determines organisational behaviour and performance. In such a setting one may ask for the qualities a leader has to have for him or her to be able to influence the organisational culture.

3.3 Theories of learning

Learning is one of the most important human traits. A driving force in learning is most likely an aspiration of people to make sense of their environment. This involves observations, memorising and learning by experimenting. One important component in learning new behaviour is also coupled to the imitating behaviour that is observed among higher primates. One could also assume that the time constants in the learning of new things was speeded up considerably when spoken language started to serve as the media for transmitting information between generations.

Individual learning has been studied for a long time by prominent scholars and many distinct theories have been presented. A consideration of these theories however suggest that they more express different situations and different views, which means that they should be combined to arrive at a model that can explain a larger set of situations. One can also argue that learning is a social process, which means that organisational learning adds on to individual learning through the group dynamics between people [2].

3.4 Bounded rationality

The concept of bounded rationality was introduced by Herbert Simon to illustrate that human decision making usually does not aim at optimal solutions, but on decisions that can be considered as satisfactory for the situation in consideration. This distinction is easy to understand in considering the need for decisions in real time together with limitations in mental capacity of the decision maker.

Human decision making has its limits. Short term memory is limited and information retrieval from long term memory is slow and unreliable. Attention, which is necessary for higher mental tasks, can be considered as a scarce resource. Still human decision makers are performing astonishingly well even in complex decision tasks, provided that they have got a suitable training. Key factors for this success lie in a combination between the ability to identify the most relevant features of a situation and the use of efficient rules of thumb for selecting paths of actions.

3.5 Possibilities in applying models of human decision making

Human decision making can be supported by various methods and tools. The methods and tools typically contain more or less explicit models of the situations for which they are created. When such methods and tools are used it is important that they are used only in situations that are within the region of validity of the underlying model. This can only be ensured if the model user is fully aware of the limitations of this model.

The largest difficulty in using models of human decision making is connected to human variability. This variability also includes skills, experiences and knowledge. The utilities are multiple and they vary with time and situation. Actual decision making always involve several stakeholders, which means that they always contain both co-operative and competitive components. Uncertainty should be modelled, but this is possible only in very special cases. All this implies that the models can only be applied in a qualitative sense.

4 COMPLEX SYSTEMS

Advances in systems theory have brought many insights into the dynamic behaviour of complex systems. Already a large dimensionality of the state space makes it very difficult to predict how they will respond to certain inputs. Furthermore nonlinearities in the interactions between variables bring in unpredictable behaviour such as bifurcations, chaos and catastrophes. Finally emergent behaviour of systems consisting of independent agents has shown that the behaviour of complex systems can only be predicted in a forward mode by simulating them with a given set of parameters.

4.1 Models

A models is always a simplification and its strengths and weaknesses lay in this simplification [3]. A model can be seen as acting as an intermediary in the transfer from micro-explanations to macro-behaviour. Because of the simplifications models have only a restricted region of validity, which means that it is necessary to have different models for different purposes. If different models are used for a specific system or situation it is important that the models are consistent with each other.

On a very general level one can separate between three problems of systems theory. The first problem is the *modelling* problem, i.e. to build a model of some observed behaviour, which may be supported by experimenting with different inputs and observing the corresponding outputs. The second problem may be termed the *simulation* problem, i.e. to predict the response of a certain system to a given input. The third problem is the *control* problem, i.e. to select the input of the system to give an output that is as near as possible to some wanted output. In principle all modelling efforts are actually performed in a pursuit of the control problem.

In creating a taxonomy of models one can separate between qualitative a quantitative models. Many of the models that are used for understanding human decision making are qualitative. Quantitative models can be further categorised into static and dynamic models, where dynamic models describe behaviour that evolves with time. Dynamic models can be further subdivided with respect to their state space into models with a finite or infinite state space. In considering quantitative models it is important to recognise that there are systems that generate qualitatively different responses for very small differences in inputs or initial states.

4.2 Control of systems

Control theory has evolved to a discipline of its own. Important characteristics of systems from a control point of view are their dynamics over time in response to control inputs. A typical distinction is to separate between the two control objectives of either maintain the state of the system at a given set point or to steer it from one defined state to another. Maintaining the state of a system at a given set point is usually arranged by feedback control, where a controller is exercising a suitable control algorithm by calculating a new input based on an error signal generated from a comparison of a set point and the observed output.

In control theory *observability* and *controllability* are important concepts that set limits on possibilities to observe hidden states of a system and to control it to some given state. Observability and controllability can in the general case be assessed only for linear systems, but theoretical results can often be used also for non-linear systems in a small region around a selected point of operation.

The industrial systems today are mostly at least on the lower levels controlled by automatic control systems that are implemented with different kinds of computers. The human control of industrial systems in production and manufacturing plants is implemented using supervisory control of these lower level automatic controls. With the increasing automation that has taken place in industrial systems, it has been possible to introduce protection systems, which take rapid action if some dangerous event is challenging plant safety. The benefit of an increased automation has been an improved safety of the plants, but has also had the drawback of an increasing complexity of the systems [4].

4.3 Agents

The concept of agents has been introduced in the field of artificial intelligence. The agents are autonomous in some sense and they have sensors and actuators with which they can observe and act on their environment. The behaviour of agents is governed by a set of programmed rules. Advances in artificial intelligence can be used to give agents a very human-like behaviour, where they are able improve and learn through their interaction with the environment and other agents. The most interesting feature of autonomous agents is that communities of agents can show an emergent behaviour of large complexity even when they are governed by very simple rules for how to act [5].

4.4 Limits of modelling

In addition to the limits set by human decision making there are hard limits on problems that can be solved. Some of these limits are set simply by a proof that a certain problem does not have any solution and other by a practical impossibility to find a solution. One simple example is a proof of non-existence, which cannot be given in the general case. Proof of non-existence can however, be given in special cases, where for instance topology makes it possible to use continuity arguments.

The perhaps best known problem that is impossible to solve is the construction of a perpetual motion machine. Another well know impossibility theorem was formulated on social welfare functions by Kenneth Arrow in the 1950's. Of interest in this connection is also the theorem proved by Alan Turing, which can be interpreted to say that it is impossibility to predict how a computer program will behave without actually running it. The theorem of Kurt Gödel can be also be interpreted to set limits of modelling, because in a given axiomatic system there are always theorems that are not provable.

Problems that are practically impossible to solve are found in cryptology and in the theory of computation. Computationally difficult is also the so called inverse problem, where the task is to identify the sources that have generated a certain observable field.

5 HIGH-RISK TECHNOLOGIES

Research in high-risk technologies has over the years created a reasonable good understanding of accidents and their causes. This research has been applied in present safety-oriented organisations, in their management systems, human system interfaces, operator training etc. Unfortunately the insights gained have not been able to prevent accidents from happening. Analyses of recent accidents have however not been able to find evidence of unexpected behaviour of systems, organisations and people, which can be seen as an assurance that present models and have been able to provide explanations at least in hindsight.

5.1 Analysing risks

Risk analysis and safety engineering are methodologies that support an understanding of possible threats to safety. Present methodologies for risk analysis include both deterministic and probabilistic approaches. The analysis is usually included as a component in a loop of analysis and synthesis, where existing solutions are analysed to identify deficiencies that can be removed or controlled for the next candidate for a system solution to be synthesised.

5.1.1 Accident models

The construction of an understanding of causes of accidents has been an important factor in building safety into high-risk technologies. The accident models that have developed over time have been important in this process. Early views on why accidents occurred made the attribution to individuals that were accident-prone. A rapid development of new technologies found the causes in design flaws, but a maturing of the technologies made these attributions increasingly strained. Human errors, which were the consequence of unsuitable human system interfaces was the next commonly seen explanation. Today the predominant view is that accidents are caused by systemic flaws that on one hand introduce hidden deficiencies and on the other hand set the conditions for triggers to be initiated [6]. A recent observation is also that such deficiencies are more likely to be created in a situation, where a safety and efficiency trade-off takes place [7].

5.1.2 Incident analysis

Incident analysis is an important component in building an understanding of causes to accidents. An important component in the success of accident analysis is an openness to reveal not only the sequence of events, but also reasons for crucial decisions that were taken. Today it is well known that a search for scapegoats cannot be a constructive approach in the analysis, but it would still be important to identify flagrant omissions of duties and attempts to hide important facts.

To be successful an incident analysis has to be both broad and deep to ensure that important issues are brought to the surface. One drawback, which the analysts should try to avoid, is that each analysis to some extent is governed by some more or less explicit model. This means that you tend to find what you are looking for. Another problem is that there is no stopping criterion for the depth of an incident analysis, which means the analysis team on quite pragmatic reasons has to define what can be considered as a reasonable scope of the analysis.

5.1.3 Safety indicators

Safety indicators are often used in safety-oriented organisations to analyse past and present performance [8]. One of the topics discussed in this connection is the separation into leading and lagging indicators, where the leading indicators would have some predictive value for future performance and the lagging indicators only measure past and present performance. Not going into the discussion of the possibility to construct true leading indicators, it may be sufficient to note that all indicator systems are depending on an underlying model of safety. If this model is valid, then also the indications given by the system can be assumed to be valid.

Indicators can correctly used be valuable to give a shorthand indication of the state of a safety-oriented organisation, but the danger is that the organisation will be controlled only through the indicators. Cases, where the indicators have been manipulated, give a clear signal that no indicator system can compensate for insightful managers, who are collecting information on performance through own observations.

5.2 Safety engineering

Design methodologies for high-risk technologies have been developed for several decades since the first large scale applications of nuclear, chemical and off-shore installations. Present design methods have much in common over areas of application, although also application specific methodologies and practices exist [9].

5.2.1 Preventing accidents

The prevention of accidents follows a simple path of identifying potential risks and designing for their removal, control and mitigation. The removal of a certain risk may for example be possible through a transfer to other processes or materials. Controlling risks can be achieved by automatic control and safety systems, which ensure that critical parameters can be held within allowable limits. Mitigation provides an additional defence, where for example certain emergency measures and evacuation ensures that damages can be held within acceptable limits.

Design for safety is often seen in building a defence in depth by erecting several independent barriers to unwanted events. The defence in depth is further elaborated into design rules, which for instance require redundancy, diversity and separation. The single failure criterion and the requirement for grace time for operators serve as additional requirements, which are guiding the design.

The prevention of accidents by design has undoubtedly had a large influence on the obtained safety level that can be observed in the operation of high-risk technologies today. There are however some remaining problems. One set of problems is caused by the possibility of common cause failures (CCF), which at the same time may offset several of the independent barriers. The other problem has to do with human attitudes, because engineered safety features may introduce a false feeling that nothing serious can happen.

5.2.2 Design basis accidents

The concept of design basis accidents (DBA) is as a design methodology, which is used to ensure that certain postulated accident sequences are properly accounted for. The design basis accidents are selected to provide a kind of worst case scenarios in a certain class of events to serve as probing stones for safety. The event sequences in the selected set of design basis accidents are simulated as accurately as possible to ensure that the engineered safety systems perform in a satisfactory manner. An account of this analysis is a central part of the safety case, which is the set of documented arguments that the design fulfils the safety requirements.

The DBA approach has the benefit of ensuring a reasonable completeness in the safety precautions, but there may still be certain sequences of events that have not been accounted for. This was for example the case with the Three Mile Island accident, where the small break loss of coolant events were assumed to be covered in the large break loss of coolant accident scenarios.

5.2.3 Human system interactions

The Three Mile Island accident pointed to the importance to consider human factors in control room design, instructions and operator training. This point is well taken today and there is a large amount of guidance available for the design of the interfaces between humans and technical systems. Instructions for disturbed and emergency operations are today also an area that is well under control through thorough testing and validation effort using simulators. Full scope simulators for the training of control room operators are today used to give the operator proficiency to handle a large set of abnormal situations.

In spite of all efforts in trying to adapt the systems to their users, the basic problem of balancing between simplicity in the design and complicated operator support systems has still to be resolved. More generally this problem has to do with the need on one hand to give the operators and decision makers the best support for a range of situations in which they may be caught and on the other hand to create flexibility in the support system to help them to respond to the situation as it is. Another of the human system interactions that still waits for its resolution is how the management system in organisations should be adapted to people [10].

5.2.4 The safety case

At least in the nuclear field it is not enough that the plants are safe, but it is in addition necessary to provide evidence to a regulator that the plants are safe [11]. The scope of a typical safety case has been increasing considerably over the years and it covers now not only the technical solutions, but also organisational provisions and staff proficiency. The safety case provides the basis for awarding an operational license and it also contains a set of conditions that have to be fulfilled during operations.

The need for a vendor and a user of a high-risk technology to prove for a third party that it can be operated safely has the benefit of initiating second thoughts and therefore also better designs. Regulatory intervention has however to rely on legislation and there is always the possibility that outdated legislation does only enforce formal safety and not real safety. Regulatory actions may also in harsh situations put an extra burden on the organisation that will redirect remedial activities to less important problems.

5.3 Safety-oriented organisations

Safety-oriented organisations have been used as a term to characterise organisations that due to the technologies they are using or the accident potential due to accumulation of materials and/or people should be focused on safety. These organisations include nuclear power, off-shore gas and oil production, transportation, health care, etc. Safety-oriented organisations have much in common, but there are also specific concerns in each of the areas mentioned. The discussion below is intended to illustrate some of the most typical precautions that are taken within these fields.

5.3.1 Authority and responsibility

The division of labour, which in organisations gives the benefit of a co-ordinated co-operation between people, presumes the definition of authority and responsibility. The most common organisational structure is the hierarchical organisation that has been used for example in the military for ages. More recent organisational innovations are the project and matrix organisations as well as process oriented work activities. Regardless of the structure however, the important thing is to assign authority and responsibility to its members. Authority and responsibility gives the members of the organisation roles in which they act. The assignment of authority and responsibility also serves to ensure commitment to organisational goals.

In assigning authority and responsibility to individuals it important also to note that these may change depending on the situation within the organisation. For example the plant may be in normal operation, it may be down for maintenance or it may experience a disturbance. The only way to ensure a safe and efficient handling of all these situations is to plan in beforehand for the situations through exercises and simulations.

5.3.2 Safety classification

One of the important principles in safety-oriented organisations is to take a graded approach to safety [12]. This means in principle that more organisational efforts are spent on systems and activities that are important for safety than on other systems and activities. This principle has lead to the so called safety classifications of systems, structures and components. The principle has been used also to some extent for work activities by defining for example certain competency requirements for specific positions.

The principle of a graded approach to safety is straightforward, but the problem is to make apply the principle in practice. In the United States the nuclear industry is only using only two safety categories, IAEA is proposing three and IEC is defining four categories. The number of categories is not important, because they could be seen as arbitrary and based on an underlying continuous safety model. Probabilistic methods have been proposed to be used to propose realistic safety classifications.

5.3.3 Safety management system

The safety management system of an organisation can loosely be seen as the part of the formal management system that oversees activities that are important for safety. Present views on managements systems are that they should integrate all organisational concerns such as safety (nuclear, environmental, labour), security (physical, information, computer), economic and judicial concerns. The management systems can be seen as the formal and described part of the organisational practices.

One important consideration in building a management system is the position that should be given the activities that are important for safety. What kind of specialised roles are found, how are the organisational structures defined and how are the planning and follow-up activities carried out? There should be instructions for the most common situations that are expected and the instructions should be assessed and updated at regular intervals.

5.3.4 Safety culture

Safety culture was a concept was introduced by IAEA after the Chernobyl accident and it has been followed up with many guiding documents. More generally safety culture could be seen as the organisational culture of a safety-oriented organisation and thus the informal fabric that makes it possible for a safety-oriented organisation to act in purposeful ways in questions that are important for safety. In an attempt to build an understanding of organisational culture, a three level models of artefacts, espoused values and hidden assumptions has been proposed [13].

The concept of safety culture has stirred a lot of discussions. Attempts have been made to measure the safety culture of an organisation using specially designed questionnaires, but even results collected with the best questionnaires leave the question open what they actually measure. It may actually be a better option not to try to assess the safety culture of an organisation and to use the concept more to initiate internal discussions to support an understanding of the safety importance of work, which is done in different parts of the organisation [14].

6 TOWARDS A SYNTHESIS

The emerging question is how all these concepts and models could be integrated to provide a better understanding for managers that are responsible for decisions that may have an influence on safety. One may also ask if there are needs for additional concepts and models, which could further improve the decisions made. However, already present models seem to provide an immense flexibility and complexity, which would suggest that a search for solutions should be started among possibilities for a better integration of present models.

6.1 Decisions in reality

The most important question in a synthesis of a decision frame, which can improve the decisions managers make in high-risk technologies is to understand what type of decisions they make. Decisions in safety-oriented organisations can be divided into two broad classes, operational and strategic decisions. This separation is somewhat fuzzy, but it basically separates between decisions that are strongly coupled to real time and decisions with no immediate time restrictions. Operational decisions should always be based on written instructions that are followed by the letter. Strategic decisions can be supported by instructions, which then mostly have a position as providing guidance.

6.1.1 Creating decision support

The creation of instructions to be used in operational decisions can be seen as design work that is initiated by strategic decisions. Basically this means that the instructions are composed of a number of decision situations in which some pre-thought sequence of actions has to be carried out. The Three Mile Island accident brought the so called n+1 instructions syndrome to the attention of human factors specialists, which implies that instructions to be followed by the letter have to include flexibility to cope with a variety of situations. Present symptom-oriented instructions seem to solve that demand fairly well.

Strategic decisions in the sense introduced above will always rely on team work with an involvement of experts from several disciplines. A typical development of a decision basis for important strategic decisions is to form a small project, which is given the necessary resources and a specified time to produce a report giving the basis for the decision to be taken. The work in the team proceeds through meetings for discussing important aspects and issues and individual work in searching for information and preparing draft sections for the report. The most important decision the team makes in this preparation is that the report is ready to be submitted to its customer. The report typically contains recommendations behind which the team stands.

6.1.2 Breadth and depth in decisions

In many decision situations it may be of help to consider two dimensions of breadth and depth in the models. The breadth has to do with the subsystems that have to be included to make the model realistic and the depth with the degree of details that has to be included in the model. A simple decision rule for the managers is to include relevant disciplines in the evaluation team when breadth is the key issue and proper specialists when the depth is the key issue. The difficulty in this connection is to be able to identify and understand the underlying issues to be able to call in appropriate resources for compiling the basis for the decision.

6.1.3 The ultimate decision maker

A typical saying in safety-oriented organisations is that everyone is responsible for safety. This is a truth that has to be qualified. It is certainly valid in the sense that the quality of all work may have direct or indirect influence on safety. Omissions in the reporting of detected anomalies may also have a safety influence. On a more general scale it is however clear that the work done in different organisational positions does not all have the same importance for safety.

A common arrangement is that one person in each organisational unit is given the authority to make decisions in the name of that unit and the responsibility for the decisions made. This person may be the CEO of a company, project manager of a project or the team leader for a small team. This responsibility is however restricted to the organisational unit and it implies that a superior unit is able to change this decision. This person can be seen as the ultimate decision maker and is therefore the main user of the decision models that are created.

This ultimate decision maker should have a good understanding of the issues for which he or she is supposed to take a stand on. This means that this person should have the necessary education and training for that position. However, it does not imply that the subordinates in an organisational unit should remain silent if they are considering some decision as a threat to safety, on the contrary they should always be encouraged to report their concerns to the next level in the organisation.

6.1.4 The need to restrict the scope

A common recommendation in decision making is to consider all influencing facts before the decision is made. This recommendation is simply impossible to fulfil. Contrary to considering all influencing facts it is necessary to concentrate on the most important issues and ensure that investigations are made with a correct balance between breadth and depth. From the point of the ultimate decision maker it also means that he or she places trust in the material that is generated. The questions to be asked are therefore if the presented evidence is complete, if there are conflicting views in the material, if is it necessary to make the decision now or if it is possible to collect more information, etc.

In giving guidance to the ultimate decision maker, one recommendation is to assess if the decision has been evaluated broadly enough. This means for example to make a check that stakeholders including both the own organisation and outside organisations have been properly accounted for. It is also important to asses how members of the own organisation will receive the decision both in a short and in a longer term. A decision is always irreversible, because when a decision has been made it will have its own influence even if this decision is reversed.

6.1.5 The validity of decision models

A decision model, which aims at some comprehensiveness, will always contain many parameters that cannot be estimated or measured. This does not necessarily make it unusable, because behaviour can always be simulated with different sets of parameter combinations, to make a qualitative analysis to identify the most important issues to be considered. More formally however, a decision model cannot usually be validated as a true description of some decision making situation.

The decision models have their most important application in preparing for predicted and possible situations, because they usually need expert involvement and plenty of time. This means that the models would be used *ex ante* for example to prepare instructions and plans for future activities. The models can naturally also be used *ex post* in incident investigations and already the large span of available models will most certainly provide models that fit almost any sequence of events to be analysed.

6.2 A set of balances

There are certain balances that become important in a strategic consideration of decisions. Such balancing dimensions can sometimes be seen as introducing opposite views, although the issue mostly is to find a suitable compromise between them. The issues considered below should be seen more as examples of important balances than as a comprehensive list.

6.2.1 Cost and quality

Cost and quality are two attributes that will be addressed in all decisions. A better quality can be obtained at some additional cost, but cost alone is not an indicator of quality. In practical situations there is some minimum requirement on quality, but it is seldom possible to give an explicit trade-off between costs and benefits of added quality. In some very specific cases it may be possible to model this trade-off to select for instance an optimal solution to a specific problem. The commonly used method to balance cost and quality issue is to use some suitably robust decision making procedure, which avoids a too large focus on only one of these two attributes.

6.2.2 Ambitions and resources

Ambition is an important criterion for decision makers in the decision processes. It is important to pursue goals that are challenging, because this may stimulate decision makers to search for further improvements to known solutions. The ambitions should however in practice always be adapted to available resources and time. The need for being realistic is also connected to varying work load that is experienced over time, which may imply that periods of large work loads should be followed by periods of smaller loads to make it possible for the organisation to recover. The discussion of resilience engineering suggests that this ability to recover is an important characteristic that should be built into the systems.

6.2.3 Competition and co-operation

Safety-oriented organisations have to build on co-operation, but competition is very ingrained in human behaviour and has therefore to be understood in decisions that are made. The balance between co-operation and competition has also to do with a sense of fairness, which seems to be inbuilt in people. Decisions that are considered to be fair have a larger possibility to be accepted and are therefore due to stimulate co-operation. Competition in the distribution of rewards is often used in organisations to motivate actors to higher performance. This practice can be very efficient, but it contains the danger of sub-optimisation and an impaired co-operation.

6.2.4 Speech or silence

Safety-oriented organisations may object to the old saying that speech is silver and silence is gold, because one key to safety is the reporting of anomalies. This may even take the form of a need to tattle on colleagues, which is considered as an offence in most cultures. However, if the motives for this reporting are clearly expressed and actual cases are handled with tact and discretion, it should be possible to foster a true reporting culture. The decision by individuals either to speak up or to remain silent is influenced by different factors such as the organisational position of the person, the own attitudes on the issue in consideration and earlier experiences from similar situations.

6.2.5 Chaos and order

The borderline between chaos and order has shown to be an interesting area of research [15]. In systems science this dichotomy has been addressed in the characterisation of communities of independent agents and in the management sciences as an explanation of innovative behaviour in small companies. A simple account of some of the research is to say that organisations need some degree of order to be functional. On the other hand too formal management methods may stifle innovation. For safety-oriented organisations the focus is undoubtedly on order, but this should not exclude an acceptance of flexibility to support organisational renewal and learning.

6.2.6 Gain and loss

Humans are typically risk averse, which means that they have a tendency to avoid gambling in situations, where they may loose. In high-risk organisation risk adverse attitudes are certainly correct, because indirect losses may easily be larger than direct losses encountered in incidents and accidents. Risk averseness may however also lead to a willingness to gamble on decisions, where over-optimistic expectations on gains are combined with miscalculated risk estimates.

6.2.7 Evolution and revolution

Organisations are supposed to develop over time. Such development is mostly evolutionary, but development may also bring the organisation to a cul-de-sac, where radical changes become necessary. The selection of an evolutionary or revolutionary path of development can be a necessary decision in organisations at some point of time. Revolutions are always expensive in terms of the uncertainty they generate and it would therefore be important to ensure that enough flexibility is retained to make it possible for the organisation to stay at an evolutionary path. On the other hand it would also be important to identify situations, where revolutions are necessary not prolong inferior performance.

6.3 A frame for understanding

The perhaps most important component of decisions within high-risk technologies is to characterise the situation correctly. What are the stakes, the possible decision alternatives and the likely outcomes? This implies a finding simplifications and the definition of a suitable decision frame. The timing of the decision is also an important component to enable the trade off between a more accurate decision basis and costs of postponing the decision.

6.3.1 Self-reflection

A decision making situation will always be determined not only by the situation, but also by the characteristics of the decision maker. This creates the need for the decision maker to include a model of him- or herself into the model of the situation. Shortcomings of knowledge in specialised fields may for instance need strengthening by suitable expert support. The assessment of the situation is a whole and ensuring a risk-neutral approach with realistic estimates of costs and benefits with selected decision alternatives may also need special considerations.

Self-reflection may also include knowledge and understanding of typical decision errors that are made together with necessary signs for their identification. Organisational drift can for example be caused by deficiencies in self-reflection. The absence of self-reflection may actually be the cause for the normalisation of deviance that was mentioned as one of the underlying causes of the Challenger and Columbia accidents [16].

6.3.2 A super-rational model

There are, as has been discussed above, many shortcomings in human and organisational decision making. A reasonable question is therefore what the possibilities are to approach the problem of decision making in complex systems. One path for a solution is to simplify, i.e. use models that hide the complexity, but still bring the essence of the situation into the foreground. This solution is straightforward and it is used in safety-oriented organisations. The problem however is that the decision model to use depends on the situation and the original decision making problem thus becomes a problem of finding a good decision model.

Another path for a solution is to build a decision making system with an emergent complexity that is able to cope with a larger set of decision situations. In such a model one could see the system as composed of many intelligent autonomous agents, who respond to their neighbours and their perceived environment with their own decision rules. This solution could actually be seen as a model of safety-oriented organisations with its components of human actors and agents realised with computerised automation. The problem however with this solution, is that the responses and the evolution of such a system are unpredictable.

6.3.3 A path towards a resolution

Both paths as laid out above contain their own unacceptable components. The first one requires a preparation of models for a large number of pre-thought situations to make it possible to create suitable action strategies. The use of such a model would require a computerised support system, where the decision maker could select a situation model that fits the situation in consideration. This is possible only if the decision maker understands the system, can use it and have confidence in the situation models it is suggesting. Another possibility is also that the decision support system would be operated by a cadre of analysts, but the problem of understanding and trust still remains.

The second path as suggested above cannot offer predictability. On the other hand the present models that are used in safety-oriented organisations cannot either provide a complete predictability. Considering a safety-oriented organisation in all its complexity as an interconnected system of thousands of autonomous intelligent agents some partial predictability may still be able to claim. It may for example be possible to argue that a certain protective action has a very high likelihood of being successful in a specific situation, because the first defence lies in a highly reliable automatic system and the second line of defence in well drilled operator actions.

6.3.4 A metaphor

Risk research has pointed to the need to include the whole society when the behaviour of high-risk technologies is modelled. This would actually suggest that the only class of models that can be used are the intelligent autonomous agents (IAA). An attempt to bring such a model to any degree of detail is not realistic, but such a model can be used in a metaphorical sense to draw important conclusions. Firstly the model cannot be used for any kind of predictions, but it may be used as a model for understanding behaviour *ex post*. Secondly it is important to give the agents behavioural strategies that are fitted to the purpose of the organisation. Thirdly there will always be conflicts between the agents that have to be resolved locally by giving some of the agent's power to enforce their own actions.

Using the IAA metaphor some general recommendations for safety-oriented organisations could be given. Managers and co-workers should be given a broad and deep background in all components pertaining to safety and decision making, because this will help them to correctly identify situations, where there is a need for something more than simple pre-thought responses to the situations. The managers at all levels should be given the integrity to withstand pressures of converting an increased safety into a higher efficiency. For the staff in safety-oriented organisations it would be advantageous if they could circulate through different positions in the organisation to be able to see the apparent conflict between safety and efficiency from different points of view.

6.3.5 Remaining research questions

The IAA metaphor could be used to investigate in greater detail its implications on safety-oriented organisations. This could for example include research to identify skills of successful decision makers. How do they formulate their own decision frames? Are there some specific rules of thumb that are helpful? What kind of strategies can be used to support decision making in groups? What kind of decision processes can be used to ensure a broad and deep consideration of safety issues?

More generally it would be important to take a look into the future. It is likely that future plants will be even more complex as compared with the plants we see today. It is also likely that the public outcry to accidents will stay at least equal to what we have today. Managers are also supposed master a broader range of disciplines and an important skill will be to motivate staff in their own organisations. With this given, what kind of organisations should we use in the future? Is it possible to build and use computerised counsellors in decision making? These are some of the possible research questions that may be addressed by future scholars in safety-oriented organisations.

7 CONCLUSIONS

One solution in the creation of better decision models for the managers in safety-oriented organisations is to have a larger variety of pre-thought situations and models that are adapted to these situations. The problem with this solution is that the search for the model to apply can be more complicated than the original decision making problem. A possible remedy is to have a decision support staff that can provide the necessary support, but the responsibility for the decision still will lie on the ultimate decision maker of the organisational unit in consideration.

Another solution in the creation of better decision models could be to use the IAA metaphor to create an organisational structure that can react on local problems with instant remedies.

The problem with this solution is that it is expected to show emergent behaviour that cannot be predicted. The demand for a covering safety case can therefore not be met, but this solution may still be acceptable and it would actually mimic the arrangement within safety-oriented organisations, where all co-workers are given the responsibility to react when they detect some anomaly.

There is no magic silver bullet that can provide better decision models for managers in safety-oriented organisations. The only solution that seems to be on hand is to do what is already done today, but to do more of it and to do it systematically with enough devotion. To what extent this work will be exploited in a higher safety or a higher efficiency is then according to the model of risk homeostasis a matter of societal choice [17].

REFERENCES

[1] Jens Rasmussen (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, SMC 13, 257-266.

[2] M.Easterby-Smith, J. Burgoyne, L. Araujo (eds.) (1999). Organizational learning and the learning organization, SAGE Publication Ltd, London.

[3] Björn Wahlström (1994). Models, modelling and modellers; an application to risk analysis, European Journal of Operations Research (EJOR), Vol.75, Issue 2.

[4] Lisanne Bainbridge (1983). Ironies of automation, Automatica, 19, 775-779.

[5] Rosario Conte, Rainer Hegselmann, Pietro Terna (eds.) (1997). Simulating Social Phenomena, Lecture Notes in Economics and Mathematical Systems 456, Springer-Verlag, Berlin.

[6] James Reason (1997). Managing the Risks of Organizational Accidents, Ashgate, Burlington, VT, USA.

[7] Hollnagel, E. (2004). Barrier analysis and accident prevention. Aldershot, UK: Ashgate.

[8] IAEA (2000). Operational safety performance indicators for nuclear power plants, TECDOC-1141, Vienna.

[9] Andrew Hale, Barry Kirwan, Urban Kjellén (2007). Safe by design: where are we now? Safety Science, vol.45, pp.305-327.

[10] Carl Rollenhagen, Björn Wahlström (2007). Management systems and safety culture; reflections and suggestions for research, Joint 8th Annual IEEE Conference on Human Factors and Power Plants and 13th Annual Workshop on Human Performance / Root Cause / Trending / Operating Experience / Self Assessment, August 26-31, Monterey, CA.

[11] Wahlström, B. (2007). Reflections on regulatory oversight of nuclear power plants, Int. J. Nuclear Law, Vol. 1, No. 4, pp.344–377.

[12] IAEA (2006). Application of the management system for facilities and activities – Safety Guide, IAEA Safety Standards Series No. GS-G-3.1.

[13] Schein E. H. (1992). Organizational culture and leadership, Jossey Bass, San Francisco.

[14] Björn Wahlström, Carl Rollenhagen (2004). Issues of safety culture; reflections from the LearnSafe project, Forth American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls and Human-Machine Interface Technologies (NPIC&HMIT 2004), Columbus, Ohio, September.

[15] M. Mitchell Waldrop (1992). Complexity; the emerging science at the edge of order and chaos, Touchstone, New York

[16] Columbia accident investigation board (2003). Report, volume 1, August, Government Printing Office, Washington DC.

[17] Gerald J.S. Wilde (1994). Target Risk, PDE Publications.