The changing nature of risk

Erik Hollnagel

The Changing Nature Of Risks

Erik Hollnagel
Ecole des Mines de Paris, Sophia Antipolis, France

Introduction

Human life has always been fraught with risks. But until the first decades of the 19th century, risks were accepted as more or less natural in the sense that they were directly associated with human activity rather than with failures of systems or equipment. Accidents happened as a part of work (which often took place at home), during major building works, when travelling on land or at sea – and of course during wars. This perception changed dramatically after September 15, 1830, when William Huskisson became the first victim of a train accident. The occasion was the opening of the Liverpool and Manchester Railway and the train that hit the unfortunate Mr. Huskisson was George Stephenson's Rocket. More accidents soon followed, involving exploding boilers, derailings, head-on collisions, collapsing bridges, and so on. (As an aside, the first recorded automobile death took place in Ireland on August 31, 1869, when a woman, Mary Ward, was thrown from and fell under the wheels of an experimental steam car built by her cousins. In 2002, road traffic accidents worldwide were estimated to kill 1.2 million people, with at least 20 million people being injured or disabled.)

The crucial change that took place in the 19th century was that accidents became associated with the technological systems that people designed, built, and used as part of work, in the name of progress and civilisation. Suddenly, accidents happened not only because the people involved, today referred to as people at the sharp end, did something wrong or because of an act of nature, but also because a human-made system failed. Furthermore, the failures were no longer simple, such as a scaffolding falling down or a wheel axle breaking. The failures were complex, in the sense that they usually defied the immediate understanding of the people at the sharp end. In short, their knowledge and competence was about how to do their work, and not about how the technology worked or functioned. Before this change happened, people could take reasonable precautions against accidents at work because they understood the tools and artefacts they used sufficiently well. After this change had happened, that was no longer the case.

The Need to Understand Risks

Risks are real in the sense that things can and do go wrong. We – society, organisations, and individuals – therefore have to deal with them. But it is important that we do this in the right way, i.e., that we understand the risks appropriately. There are many definitions of risk, but most of them involve the notion of an adverse outcome or a potential negative impact that

arises from some present process or future event. The occurrence of the event is possible rather than certain, either because it is unknown or or because it occurs with some probability. This also means that the loss is probable rather than certain. A risk is deemed to be large if either the loss is severe, if the probability is high, or both together. Similarly, a risk is deemed to be small if the loss is small, if the probability is low, or both together.

Since negative outcomes are unwanted and undesirable, everyone – individuals, organisations, and society – are interested in finding ways to avoid that these outcomes happen. For example, we all know that it is risky to drive a car in traffic or to cross a busy street, but we do not know *when* a traffic accident involving us will happen. We therefore proceed with as much caution as we find necessary to remain safe. The same goes for individuals at work and for the larger socio-technical systems. But where the individual normally can rely on common sense and experience, socio-technical systems must employ more direct and explicit methods In order for a system to avoid accidents, which under normal circumstances is tantamount to being safe, it is critical to be able to identify and manage risks, and therefore to understand what the risks are in the first place. Classical risk assessment, for instance, normally starts from the unwanted consequences, such as the top event in a fault tree. In order to do this, the unwanted consequence must be recognisable either because it has happened before, which means that it is part of the individual or joint experience, or because it can be imagined – which usually means that it is a linear extrapolation of something that has happened before.

The Difficulty in Understanding Risks

Safety can be defined as the absence of adverse outcomes (accidents, incidents, personal injuries, work loss days, etc.), or more formally as a state in which the risk of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level. This safe state is achieved through a continuing process of hazard identification and risk management. Regardless of the precise definition, a critical prerequisite for safety is the ability to identify in advance the events that may lead to adverse outcomes, as well as the outcomes themselves. This is, indeed, what risk assessment is all about, and over the years a large number of methods and techniques have been developed to make this process more efficient and reliable. Methods and techniques are, however, of limited value unless they are based on an adequate understanding of the domain and the types of work involved. No risk assessment methods can be applied in a mechanically or unthinking fashion. On the contrary, effective risk assessment depends critically on the ability of investigators and analysts to imagine what can possibly go wrong. This ability, or requisite imagination (Adamski & Westrum, 2003), comprises three steps. The first step is to understand what the problem is or indeed to appreciate that there is a problem at all. The second step is to understand the "mechanisms" or the ways in which the adverse outcomes can arise, to envisage the consequences, and to differentiate between large and small risks. The third and final step is to think of or find the means which can be used either to reduce or eliminate the risk, or to protect against the consequences. If one or more of these steps fail, the risk may not be

noticed until something happens, at which time it is usually too late to do anything about it. Two characteristic examples will hopefully make clear what the three steps mean.

An uncomplicated risk: Smoking and cancer

As an example of a risk that is relative easy to comprehend, even for non-specialists, consider the relation between smoking and lung cancer. Ever since the publication of the British doctors study in the 1950s, it has been common knowledge – except, perhaps, in the tobacco industry – that tobacco smoking increases the risk of lung cancer. (But notice that before this study, few people considered smoking a risk. The study established an irrefutable statistical and causal relation between smoking and lung cancer.) The way in which this happens, the "mechanism", is well described and well understood; it is easy to envisage the consequences and therefore to differentiate between large and small risks, for instance between active and passive smoking. Finally, the solution to the problem is also known and in itself quite uncomplicated, although it sometimes seem to be difficult to apply individually. (This therefore nicely illustrates that knowing that a risks exists is a necessary but not a sufficient condition for reducing or eliminating it.) The relation between smoking and cancer is nevertheless a risk that is easy to understand.

A complicated risk: Global warming

We can use another kind of "smoking" as an example of a risk that it is difficult to understand. Global warming, also known as the greenhouse effect, is the phenomenon that changes in the levels of carbon dioxide in the atmosphere can lead to changes in the surface temperature of our planet. Although it is the general consensus of scientists and experts, such as the Intergovernmental Panel on Climate Change, that global warming is a reality, it nevertheless remains a fiercely debated issue. There are still many people, well-known writers, scientists, and politicians among them, that flatly deny the existence of global warming. In terms of the three steps mentioned above, already the first seems to be hard, i.e., it seems to be difficult to acknowledge that there is a problem at all. (This may, of course, be due to other reasons, such as economic interests and political expediency; the problem may thus be understood, but despite that not acknowledged.) So while for some people the problem is real, for others it is only an environmentalist fantasy. The second step is to understand the "mechanisms" and the ability to envisage the consequences. As far as the mechanisms are concerned, they have been known since the Swedish scientist Svante Arrhenius in 1895 presented a paper to the Stockholm Physical Society entitled "On the influence of carbonic acid in the air upon the temperature of the ground." (The paper was published the following year. Yet it is a surprise to many people today that the greenhouse effect was described so long ago.) As far as the effects are concerned, estimates of their magnitude vary considerably; some even see global warming as a positive development (Arrhenius was in fact of that opinion himself). The third step is also difficult, since it is not easy to think of ways in which the risk or the outcomes can be reduced. (In this case, abandoning the collective "smoking" that leads to global warming may be even more difficult than in the case of individual smokers.) All in all, global warming is an example of a risk that is difficult to comprehend.

Large and small risks

Practically all industries explicitly deal with and acknowledge the serious risks, mostly because they understand the benefits of doing so, but sometimes simply because they have to. This has over many years established a practical understanding in the perception and handling of risks across industries and domains. One example of that is the ALARP (As Low As Reasonably Practicable) principle, where the determination of what is "reasonably practicable" reflects a combination of economic, practical, and ethical concerns.

The same does not go for risks with less spectacular outcomes. In these cases it is often difficult to understand what the problem is, or sometimes even to see that there is a problem at all – at least not until something has happened. It is the irony of risk assessment that the success of eliminating the large problems, where the "mechanisms" are easy to understand, inevitably and unfortunately leaves the problems that are harder to understand. Adverse outcomes are not always due to cause-effect chains or a linear propagation of the effects of a malfunction, but may also arise from unusual combinations of conditions that involve poorly understood characteristics of the socio-technical systems.

If socio-technical systems were relatively stable and only changed slowly, the experience from accidents and incidents that happened would over time be sufficient to ensure an acceptable level of safety. Unfortunately, industrialised societies continue to develop and the socio-technical systems become ever more complex. This means that the risks also change and that accumulated experience never will be sufficient. Since risk assessment and accident analysis methods necessarily are a product of accumulated experience, there will unfortunately and invariably be a lag between the changes in the real world and the corresponding changes or renewals or updates of models and methods. In other words, even if the risks of a system have been fully understood at one point in time (and even that may be debatable), this will not be sufficient to guarantee a safe state in the future.

The growing complexity of socio-technical systems

One useful characterisation – if not quite an explanation – of this development was given by the American sociologist Charles Perrow in a book called Normal Accidents (Perrow, 1984). The fundamental thesis of the book was that the industrialised societies, and in particular the technological environments that provided the foundation for those societies, by the end of the 1970s had become so complex that accidents were bound to occur. Accidents were thus an inevitable part of using and working with complex systems, hence should be considered as normal rather than rare occurrences. Since Perrow published his analyses neither the socio-technical systems, nor the problems that follow, have become any simpler.

Perrow built his case by going through a massive set of evidence from various types of accidents and disasters. The areas included were nuclear power plants, petrochemical plants, aircraft and airways, marine accidents, earthbound systems (such as dams, quakes, mines, and lakes), and finally exotic systems (such as space, weapons and DNA). The list was quite

formidable, even in the absence of major accidents that occurred later, such as Challenger, Chernobyl, and Zebrügge.

Perrow proposed two dimensions to characterise different types of accidents: *interactiveness* and *coupling*. With regard to the interactiveness, a complex system – in contrast to a linear system – was characterised by the following:

- Indirect or inferential information sources.
- Limited isolation of failed components.
- Limited substitution of supplies and materials.
- Limited understanding of some processes (associated with transformation processes).
- Many control parameters with potential interaction.
- Many common-mode connections of components not in production sequence.
- Personnel specialization limits awareness of interdependencies.
- Proximate production steps.
- Tight spacing of equipment.
- Unfamiliar or unintended feedback loops.

According to Perrow, complex systems were difficult to understand and comprehend and were furthermore unstable in the sense that the limits for safe operation (the normal performance envelope) were quite narrow. Perrow contended that we have complex systems basically because we do not know how to produce the same output by means of linear ones. And once built, we keep them because we have made ourselves dependent on them.

Systems can also be described with respect to their coupling, which can vary between being loose or tight. The meaning of coupling is that subsystems and/or components are connected or depend upon each other in a functional sense. Thus, tightly coupled systems are characterised by the following:

- Buffers and redundancies are part of the design, hence deliberate.
- Delays in processing not possible.
- Sequences are invariant.
- Substitutions of supplies, equipment, personnel is limited and anticipated in the design.
- There is little slack possible in supplies, equipment, and personnel.
- There is only one method to reach the goal.
- Tightly coupled systems are difficult to control because an event in one part of the system quickly will spread to other parts.

Perrow used these two dimensions of interactions and coupling to illustrate differences among various types of systems, cf. Figure 1.

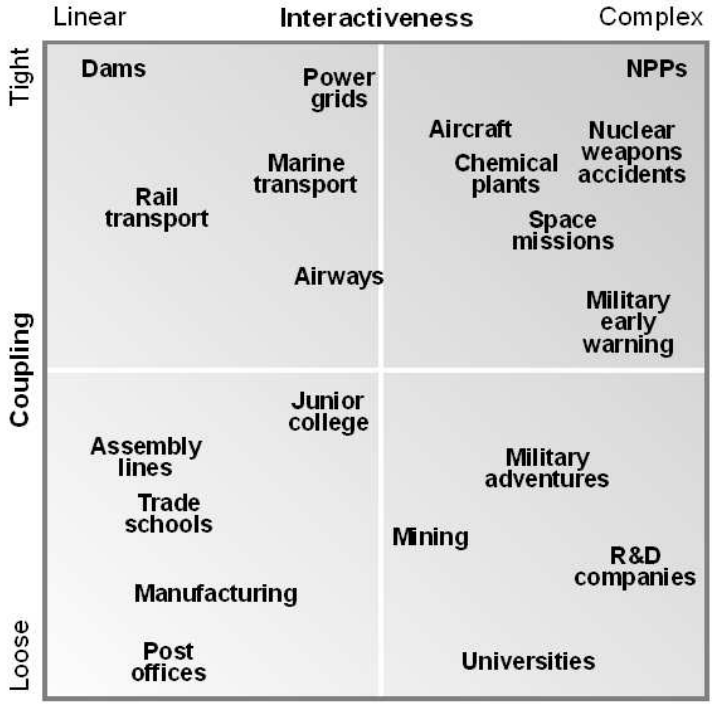

Figure 1: The coupling-interactiveness diagram (after Perrow, 1984)

The worst possible combination in terms of risk and accident potential is, of course, a complex and tightly coupled system. Perrow's prime example of that was the nuclear power plant, with Three Mile Island accident as a case in point. Other systems that belonged to the same category were, e.g., aircraft and chemical plants. It was characteristic, and probably not a coincidence, that all the systems Perrow described in the book were tightly coupled and only differed with respect to their complexity, i.e., they were mostly in the upper right quadrant.

Perrow's thesis, as expressed by Figure 1, is relevant for risk assessment methods since the understanding of risk, either in accident investigation or in risk assessment, must be able to account for the nature of interactions and the degree of coupling in the system. If we, for the sake of argument, refer to the four quadrants of Figure 1, then it is clear that systems in the lower left quadrant in important respects differ from systems in the upper right quadrant. A method that may be adequate to understand risks and adverse outcomes in a system in the lower left quadrant, such as a person being injured while working at an assembly line, is unlikely to be sufficient to explain risks and adverse outcomes in a system in the upper right quadrant, such as an event at a nuclear power plant serious enough to be rated on the International Nuclear Event Scale (INES). (Even though the converse is not necessarily true, it may be inefficient to use the more complex and powerful methods to investigate accidents or assess risks in simple systems.) The diagram therefore provides an external frame of reference for risk assessment methods in addition to the more traditional requirements such as consistency, reliability, usability, etc.

In the description proposed by Perrow (1984), the notion of coupling is relatively straightforward. But the notion of complexity must be used with some care, since it can refer either to the epistemological or the ontological complexity (Pringle, 1951), i.e., either the complexity of the description or the supposedly "true" complexity of the system. For practical reasons it is preferable to use a different concept, namely how easy it is to manage or control the system, where the extremes are tractable and intractable systems. A system, or a process, is tractable if the principles of functioning are known, if descriptions are simple and with few details, and most importantly if the system does not change while it is being described. Conversely, a system or a process is intractable if the principles of functioning are only partly known or even unknown, if descriptions are elaborate with many details, and if the system may change before the description is completed. A good example of a tractable system is the normal functioning of a post office, or the operation of a home furnace. Similarly, a good example of an intractable system is the outage at a NPP or the activities in a hospital emergency department. In the latter cases the activities are not standardised and change so rapidly that it is never possible to produce a detailed and complete description (Wears et al., 2006).

Using this modification of the terminology, we can propose a new version of Perrow's diagram, as shown in Figure 2. (Note that this also means that some of the examples used by Perrow have to change position; in addition, some examples (e.g., nuclear weapons accidents) have been deleted, while others (financial markets) have been introduced. These changes are, however, illustrative rather than exhaustive.)

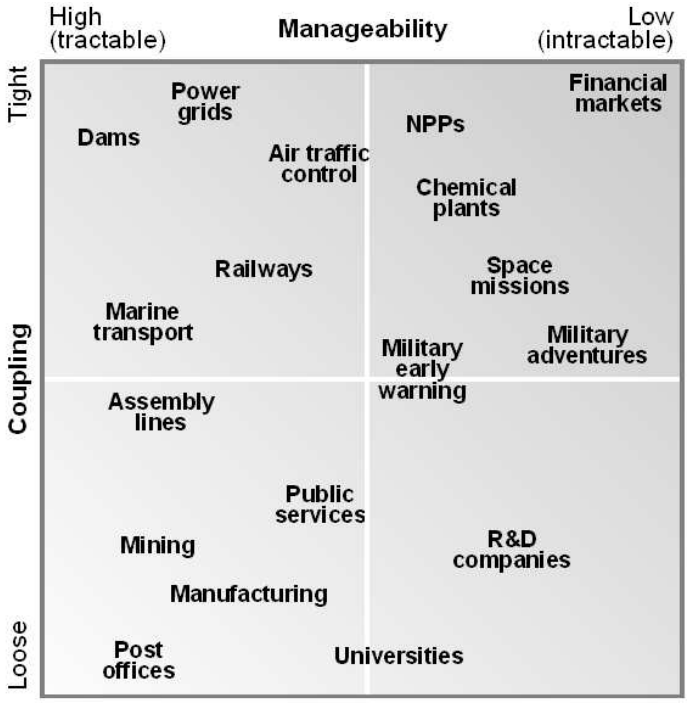

Figure 2: Revised Perrow diagram

Following this principle, risk assessment methods should be characterised in terms of the systems – or conditions – they can account for. For instance, a simple linear model – such as the domino model (Heinrich, 1931) – can be used to account for certain types of risks and not for others. The domino model is suitable for systems – hence for accidents – that are loosely coupled and tractable. This is not very surprising, since most systems were of that type at the time the domino model was developed. Nuclear power plants considered as systems are, however, tightly coupled and more or less intractable and require models and risk assessment methods that are capable of accounting for these features. It is therefore reasonable to characterise investigation methods in terms of which applications they can account for. While this will not by itself determine whether one method is "better" than another, it will make it possible to choose a method that is suitable for a specific purpose and/or system and thereby also to exclude methods that are unable to meet the requirements of a given assignment.

The Power of Risk Assessment Methods

When problems are difficult to understand, it also becomes difficult to envisage the consequences and to pinpoint the significant risks. The events that one would want to avoid may first of all only occur very infrequently, cf., Westrum's (2006) notion of irregular threats. The events may even belong to the category of rare events – meaning that they are almost never repeated. Since their aetiology defies traditional explanations or accident models, it is usually difficult both to determine what consequences may obtain and to assess their likelihood. Finally, even when the risks can be assessed, the absence of easily understandable causes makes it hard to propose concrete and cost-efficient countermeasures. Without a clear focus, it is very difficult to know how to respond.

The problem is not made easier by the ongoing change from linear to non-linear accident and safety models. This change is a consequence of the growing recognition that accidents can be due to couplings or interactions among functions or events that are not in themselves failures or malfunctions, hence are not found by traditional risk analyses. One way of expressing that is to note that accidents more often are due to usual actions under unusual circumstances than to unusual actions under usual circumstances. In other words, the explanations cannot be found nicely tucked away in a single part of a socio-technical system, such as the operator or the interface, but are rather due to the ways in which normal performance variability can combine in unexpected ways. For risk assessment this creates a need for models and methods that can explain how adverse events can arise from normal performance variability as well as from failures and malfunctions.

The description of three steps in risk assessment, acknowledging that a problem exists, understanding the "mechanisms" and differentiating various consequences, and finding effective means, is applicable to both linear and non-linear accident types. But if it is hard to understand the "mechanisms" for classical risks, it is even more challenging for the risks that are described as emerging from more complex, socio-technical systems. Yet it is essential that we become able to do that, not only on the level of analysis but also on the levels of management and policy making. A failure to do so will limit safety management to be reactive

and risk analyses to be nothing more than extensions of error counting. Yet failure is not an option, neither in what we do, nor in how we do it.

Changing notions of risk and safety

Most of the methods for risk assessment and accident investigation that are used today in safety critical industries have their origin in the 1960s. This was the period where new analysis methods were required to match the growing complexity, and therefore also the growing risk, of technological systems. Examples are Fault Trees, which were developed in 1961 to evaluate the launch control system for the Minuteman ICBM (cf. Leveson, 1995), Hazard and Operability Analysis (HAZOP) which was developed by Imperial Chemical Industries in England in the early 1960s (CISHC, 1977), and Failure Mode and Effects Analysis (FMEA) which was originally developed by the US military in 1949 but later superseded by the Failure Mode, Effects and Criticality Analysis (FMECA) (MIL-STD-1629A, 1980).

Another period of rapid growth occurred in the beginning of the 1980s, mainly in response to the TMI accident in 1979. This led to the recognition that human factors and human errors played a significant role in system safety, hence that it was necessary for risk assessment and accident investigation methods to go beyond the technological system. The concern for the human factor was later extended to cover organisations and organisational factors as well, with the prominence of 'safety culture' as a good example. The direct motivation was also in this case a serious adverse event, namely the Chernobyl accident in 1986. Since the mid-1990s there has been an additional growth, although more often incremental than innovative. This growth has taken place to meet the perceived need among researchers and practitioners of a re-orientation in thinking about safety, in order to develop methods and approaches that are both more efficient in use and better grounded in their concepts and constructs.

Some of the major changes and developments since the mid-1990s have been:

- An increasing emphasis of the organisational factor, spurred by Jim Reason's book on organisational accidents (Reason, 1997),
- the increasing importance of software (e.g., the concept of Safeware; Leveson, 1995),
- the emphasis on high reliability organisations, (e.g., Weick, Sutcliffe & Obstfeld, 1999)
- the changing perspective on causality, moving from sequential models to systemic models (Hollnagel, 2004),
- the associated change in view on "human error", from the "old" look to the "new" look (Dekker, 2006),
- the change from training in specific skills to training in general communication and collaboration (Helmreich, Merritt & Wilhelm, 1999),
- the change from reactive to proactive safety, as marked by resilience engineering, (Hollnagel, Woods & Leveson, 2006).

In the same period, i.e., since the mid-1990s, the growing complexity of socio-technical systems has also necessitated the development of more powerful accident investigation and risk assessment methods and a revision of the underlying analytical principles. This complexity, which was aptly diagnosed by Perrow (1984), has unfortunately often been marked by serious accidents, and shows no sign of abating. Some of the better known examples are the JCO accident at Tokai-Mura, Japan (1999), the space shuttle Columbia disaster (2003), and the Überlingen mid-air collision (2002) – plus literally thousands of small and large accidents in practically every industrial domain. This development has not been isolated to a specific domain but can be found in many different industries and service functions.

One consequence of this has been the realisation that accident investigation and risk assessment are two sides of the same coin, in the sense that they consider the same events or phenomena either after they have happened (retrospectively) or before they happen (prospectively). In the prospective case there is, of course, the possibility that an event may never occur; indeed, the main rationale for risk assessment is to ensure that this is the case. The dependency between accident investigation and risk assessment has been emphasised both by the so-called second generation HRA methods (in particular ATHEANA, Cooper et al., 1996; CREAM, Hollnagel 1998; and MERMOS, Le Bot, Cara & Bieder, 1999), and is also a central premise for Resilience Engineering (Hollnagel, Woods & Leveson, 2006).

Development of New Accident Analysis and Risk Assessment Methods

One reason for the development of new methods and approaches has been the inability of established methods to account for novel types of accidents and incidents. Another reason has been a lack of efficiency, in the sense that recommendations and precautions based on the usual explanations have not lead to the desired effects and improvements. A third reason has been new theoretical insights, although this rarely has happened independently of the former.

In the two cases the inability and/or lack of efficiency of existing methods is a consequence of the continued, rapid development of socio-technological systems, in turn driven by a combination of technological innovation, commercial considerations, and user demands. This contrasts with risk assessment and safety management methods that develop at a much more moderate pace – if at all – which means that they rarely are able to represent or address the actual complexity of industrial systems. To the extent that methods develop, it is usually as a delayed reflection of "new" types of accidents. The outcome can be that new methods focus on a specific, salient factor of an event (e.g., violations after Chernobyl), or that they become more comprehensive by trying to draw together the collective experience and changes in view (e.g., second generation HRA).

In order to determine whether a given method is adequate for a given system and scenario, it is necessary to be able to characterise both. A system – or a scenario – can conveniently be described using the dimensions of *coupling* and *manageability*, cf., Figure 2 above. For the

sake of this discussion, we will assume that the dimensions can be considered as binary. This leads to the following four classes of systems.

- Systems that are loosely coupled and tractable (lower left quadrant)
- Systems that are tightly coupled and tractable (upper left quadrant)
- Systems that are loosely coupled and intractable (lower right quadrant)
- Systems that are tightly coupled and intractable (upper right quadrant)

The various accident investigation and risk assessment methods can in a similar manner be characterised in terms of the assumptions they make about the nature of risks. For instance, whether risks are seen as being due to single failures and malfunctions, to human factors, to combinations of failures and weakened defences, or to systemic failures. Combining these characterisations gives rise to the following considerations.

Methods suitable for systems that are loosely coupled and tractable

In terms of frequency or numbers, most systems are even today loosely coupled and tractable. Many of the commonly used investigation methods are best suited for systems with those characteristics – or even explicitly assume that this is the case. In practical terms this implies that it must be possible to provide a more or less complete description of the system and to account for events (e.g., failures or malfunctions) in a one-by-one or element-by-element fashion. While these assumptions make for methods that are easy or simple in terms of use, it also means that such methods are inadequate for systems in high-risk domains, such as nuclear power production, chemical production, or air traffic management.

Out of the many types of methods that are adequate for loosely coupled and tractable systems, a number of characteristic subtypes can be distinguished.

Methods that focus on the identification of failed barriers

The Accident Evolution and Barrier Function (AEB; Svensson, 2001) is a method that focuses on barriers and/or defences and explains accidents as the result of failed or deficient barriers. It is primarily an accident investigation method that describes the evolution towards an accident or incident as a series of interactions between humans and technical systems. The interactions are represented as failures, malfunctions or errors that could lead to or did result in an accident. The method forces analysts to integrate human and technical systems simultaneously when performing an accident analysis.

The method starts by modelling the accident evolution in a flow diagram. The AEB method only models errors and therefore does not work with or represent the full event sequence. The flow chart initially consists of empty boxes in two parallel columns, one for the human systems and one for the technical systems. The second phase consists of the barrier function analysis. In this phase, the barrier functions are identified as the failures, malfunctions or errors that constitute the accident evolution, i.e., as error boxes. In general, the sequence of error boxes in the diagram follows the time order of events. Between each pair of successive error boxes there is a possibility to arrest the evolution towards an incident/accident. According to the AEB model, the same barrier function can be performed by different barrier

function systems. Correspondingly, a barrier function system may perform different barrier functions.

The result of an AEB analysis is a list of broken barrier functions, the reasons for why there were no barrier functions or why the existing ones failed, and to suggestions for improvements.

Methods that focus on human error

HERA (Human Error in ATM) is a an examples of a method that focus on human error as the primary contributor to risks and adverse events (Isaac, Shorrock & Kirwan, 2002). The purpose of HERA is to identify and quantify the impact of the human factor in incident/accident investigation, safety management and prediction of potential new forms of errors arising from new technology. Human error is seen as a potential weak link in the Air Traffic Management (ATM) system. Measures must therefore be taken to prevent errors and their impact, and to maximise other human qualities such as error detection and recovery. HERA is predicated on the notion that human error is the primary contributor to accidents and incidents.

The HERA method comprises the following steps:

1. Defining the error type.
2. Defining the error or rule breaking or violation behaviour through a flowchart.
3. Identifying the Error Detail through a flowchart.
4. Identifying the Error Mechanism and associated Information Processing failures through flowcharts.
5. Identifying the tasks from tables.
6. Identifying the Equipment and Information from tables.
7. Identifying all the Contextual Conditions through a flowchart and tables.

The outcome of a HERA analysis is the identification of human errors and violations, with quantitative data on the relative frequency of error types and working conditions.

Methods that focus on root causes in isolation

The purpose of root cause analysis (e.g., Wilson et al., 1993) is to identify the deficiencies in a safety management system that, if corrected, would prevent the same and similar accidents from occurring. Root cause analysis is a systematic process that uses the facts of the accident to determine the most important reasons or causes.

1. Determine sequence of events
2. Define causal factors
3. Analyse each causal factor's root causes
4. Analyse each root cause's generic causes
5. Develop and evaluate corrective actions
6. Report and implement corrective actions

The result of a root cause analysis if the identification of specific (root) causes that then can be made the object of specific remedial or corrective action.

Methods that focus on root causes in combination

Although it in some cases may be sufficient to look for and find specific causes, most industrial systems are designed so that single failures will not constitute a risk or lead to an accidents. Risks are therefore more often due to a combination of individual failures, and methods are therefore needed that can accommodate that.

One example of such methods is HINT (Takano, Sawayanagi & Kabetani, 1994), which is based on the Japanese version of the the Human Performance Enhancement System (HRES; INPO, 1989). The overall principle of HINT is to make a root cause analysis of small events to identify trends, and to use this as a basis for proactive prevention of accidents. The same principles can be found in SAFER (Yoshizawa, 1999), although the latter method has a wider scope, and therefore may be applicable to accidents in tightly coupled systems as well.

The HINT method comprises the following four steps.

1. Understand the event.
2. Collect and classify causal factor data.
3. Causal analysis, using root cause analysis.
4. Proposal of countermeasures.

The method differs from the traditional root cause analysis by focusing on minor human error events, i.e., on incidents rather than accidents. By supporting a trend analysis of these events, it becomes possible to consider safety proactively and to focus on the prevention of serious accidents.

Methods suitable for systems that are tightly coupled and tractable

The increasing frequency of non-trivial accidents during the 1980s and 1990s made it clear that explanations in terms of sequences or chains of causes and effects were insufficient. This also meant that risk assessment could not be limited to looking for single failures or malfunctions – whether of technical components or humans. In order to be able to deal with the increasingly complex systems, it was necessary to account for how combinations of multiple sequences of events, or of events and latent conditions, could arise. This led to the proposal of complex linear models, sometimes also called epidemiological models (Hollnagel, 2004). The two major types of methods suitable for tightly coupled and tractable systems are associated with the Swiss cheese model and the Man-Technology-Organisation (MTO) model. A third and principally different approach is the Cognitive Reliability and Error Analysis Method (CREAM), which also can be seen as a precursor of methods applicable to tightly coupled, intractable systems.

The Swiss cheese model (SCM)

One of the best known accident investigation methods of the 1990s is associated with the so-called Swiss Cheese model (Reason, 1990). This model represents an organization's defences against failure as a series of barriers, represented as slices of Swiss cheese. (To be precise, this

must be the Emmenthaler cheese, which is a medium-hard cheese with characteristic large holes.) The holes in the cheese slices represent weaknesses in individual parts of the system that are assumed to vary continually in size and position in the slices. The holes can therefore also be seen as representing the risks in a system. According to this analogy, an accident can happen when holes in each of the slices momentarily align, permitting "a trajectory of accident opportunity", so that a hazard passes through all of the holes in all of the defenses, leading to a failure.

The basic method for using the SCM is to trace backwards from the accident. The analysis looks for two main phenomena: *active failures*, which are the unsafe acts committed by people (slips, lapses, fumbles, mistakes, and procedural violations); and *latent conditions*, which arise from decisions made by designers, builders, procedure writers, and top level management. Latent conditions can translate into error provoking conditions within the local workplace and they can create long-lasting holes or weaknesses in the defences. Unlike active failures, whose specific forms are often hard to foresee, latent conditions can be identified and remedied before an adverse event occurs. Understanding this can support proactive rather than reactive risk management. There are several specific methodologies associated to the Swiss cheese model, the best known being the TRIPOD method (Hudson, Primrose & Edwards, 1994).

MTO (Människa-Teknologi-Organisation or Man-Technology-Organisation)

Another method is the so-called MTO-analysis, which explicitly considers how human, organisational, and technical factors can interact to constitute a risk, and therefore also serve to explain accidents that have happened (Bento, 1992; Rollenhagen, 1995). An MTO investigation comprises three methods:

1. Structured analysis by use of an event- and cause-diagram.
2. Change analysis by describing how events have deviated from earlier events or common practice.
3. Barrier analysis by identifying technological and administrative barriers which have failed or are missing.

The first step in an MTO-analysis is to develop the event sequence longitudinally and illustrate the event sequence in a block diagram. Then, to identify possible technical and human causes of each event and draw these vertically to the events in the diagram, i.e., as factors or conditions influencing the event. The next step is to make a change analysis, i.e. to assess how events in the accident progress have deviated from normal situation, or common practice. Further, to analyse which technical, human or organisational barriers have failed or were missing during the accident progress. The basic questions in the analysis are how the continuation of the accident sequence could have been prevented, and what the organisation could have done in the past in order to prevent the accident.

The last step in the MTO-analysis is to identify and present recommendations. These should be as realistic and specific as possible, and might be technical, human or organisational. The MTO analysis thus produces a detailed description and a clarification of factors that either led to or contributed to the accident.

Cognitive Reliability and Error Assessment Method (CREAM)

CREAM was developed to be used both predictively and retrospectively (Hollnagel, 1998). Unlike the Swiss cheese and the MTO approaches, CREAM has a clearly defined theoretical basis in the Contextual Control Model (COCOM). This emphasises that risks are a function of the degree of control in a socio-technical system, and associates the degree of control with four different modes called strategic, tactical, opportunistic, scrambled, respectively. It is assumed that a lower degree of control corresponds to less reliable performance. The level of control is mainly determined by the Common Performance Conditions (CPC), i.e., by external factors rather than by internal failure probabilities. The retrospective use of CREAM (accident investigation) is based on a clear distinction between what can be observed (called phenotypes) and what must be inferred (called genotypes). The genotypes used in CREAM are divided into three categories: individual, technological and organisational, corresponding to the MTO triplet.

The procedure for CREAM for accident investigation comprises the following steps:

1. Produce a description of what actually happened
2. Characterise Common Performance conditions
3. Produce a time-line description of significant events
4. Select all actions of interest
5. For each action, identify failure mode (this is done iteratively)
6. For each failure mode, find relevant antecedent-consequent links (this is done recursively)
7. Provide overall description and draw conclusions.

The analysis can be documented by a graph, or a network, of antecedent actions (functions) and conditions that together constitute an effective explanation of the accident. The graph shows how various actions and conditions affected each other in the given situation. The use of CREAM for risk assessment basically follows the same approach, leading to a value for the failure probability (Fujita & Hollnagel, 2004).

Methods suitable for systems that are loosely coupled and intractable

There are no methods applicable to socio-technical systems in this category. The reason for that has to do with the historical development of accident investigation and risk assessment methods. At the beginning, effectively in the 1930s, industrial systems were loosely coupled and tractable. As technologies and societies developed, systems became more tightly coupled through vertical and horisontal integration, and at the same time less tractable because new technologies allowed faster operations and more extensive automation. The latter meant in particular that they became more or less self-regulating under normal conditions, which reduced tractability. Since accidents 'followed' these developments, methods were developed to be able to adress the new problems. Conversely, few if any accident of note took place in loosely coupled, intractable systems, hence no methods were developed to account for that. The basic reason is that such systems are social rather than technological, e.g., universities,

research companies, and the like. They are therefore not designed in the same sense, nor do they have the potential for accidents with direct consequences for human life and/or material.

Methods suitable for systems that are tightly coupled and intractable

The continuously growing complexity of socio-technical systems, and the consequent reduction of tractability, has led to a fundamental change in the approach to risk and safety. The most prominent example of that is the development resilience engineering (Hollnagel, Woods & Leveson, 2006), which changes the focus from failures and actions gone wrong to the usefulness of normal performance variability. With respect to accident investigations this means that the aim is to understand how adverse events can be the result of unexpected combinations of variations in normal performance, thereby avoiding the need to look for a human error or a root cause. This view is often referred to as a systemic view. There are presently two main proposals for a method, STAMP and FRAM.

System-theoretic model of accidents (STAMP)

The hypothesis underlying STAMP is that system theory is a useful way to analyze accidents, particularly system accidents (Leveson, 2004). Accidents occur when external disturbances, component failures, or dysfunctional interactions among system components are not adequately handled by the control system. Safety is viewed as a control problem, and is managed via constraints by a control structure embedded in an adaptive socio-technical system. Understanding why an accident occurred requires determining why the control structure was ineffective. Preventing future accidents requires designing a control structure that will enforce the necessary constraints. Systems are viewed as interrelated components that are kept in a state of dynamic equilibrium by feedback loops of information and control. STAMP thus uses a feedback control system as a specific causal model. A STAMP analysis proceeds along the following lines:

1. In teleological systems, various subsystems maintain constraints which prevent accidents.
2. If an accident has occurred, these constraints have been violated.
3. STAMP investigates the systems involved, especially human-organisational subsystems, to identify missing or inappropriate features (those which fail to maintain the constraints).
4. It proceeds through analysing feedback & control operations.

The most basic component of STAMP is not an event, but a constraint. Risks and accidents are therefore viewed as resulting from interactions among components that violate the system safety constraints. The control processes that enforce these constraints must limit system behavior to the safe changes and adaptations implied by the constraints. Inadequate control may result from missing safety constraints, inadequately communicated constraints, or from constraints that are not enforced correctly at a lower level.

Functional Resonance Accident Model (FRAM)

If it is acknowledged that risks and accidents can arise from unexpected combinations of normal performance variability, then the assumption of causality must be partly abandoned. If risks and accidents cannot always be linked to failures and malfunctions of components, then methods should not be restricted to causal explanations. The alternative is to develop methods for accident investigation and risk assessment that describe system functions rather than components or structures, and that can account for the non-linear propagation of events. This can, for instance, be achieved by using functional resonance instead of causality, and by using normal performance variability instead of malfunctioning (e.g., Hollnagel, 2004; Sawaragi, Horiguchi & Hina, 2006).

The method associated with FRAM proceeds along the following steps:

1. Define the purpose of modelling and describe the situation being analysed. The purpose can be either risk assessment of accident investigation.
2. Identify essential system functions and characterise each function by six basic parameters (input, output, time, control, pre-conditions, resources).
3. Characterise the (context dependent) potential variability using a checklist. Consider both normal and worst case variability.
4. Define functional resonance based on possible dependencies (couplings) among functions.
5. Identify barriers for variability (damping factors) and specify required performance monitoring.

The analysis uncovers dependencies among functions or tasks that normally are missed. It also identifies the information needed for the investigation. The concrete result can be a graphical rendering of how the accident developed and/or a detailed written description (Nouvel, Travadel & Hollnagel, 2007). The basis for a risk assment is the performance variability of normal actions.

Discussion and Conclusions

One way of summarising the characterisation of the methods described above is to map them onto the diagram shown in Figure 2. The result of that is can be seen in Figure 3. This shows that most methods are applicable to tractable systems, or rather that most methods assume that the systems are tractable. Conversely, one may conclude that these methods should not be used for intractable systems, since they will not be able to produce adequate explanations. Several of the commonly used methods, including root cause analysis, AEB, and HERA, also require that systems only are loosely coupled. These methods are therefore unable to account for the consequences of tight couplings, hence unable adequately to explain accidents in systems of that type.

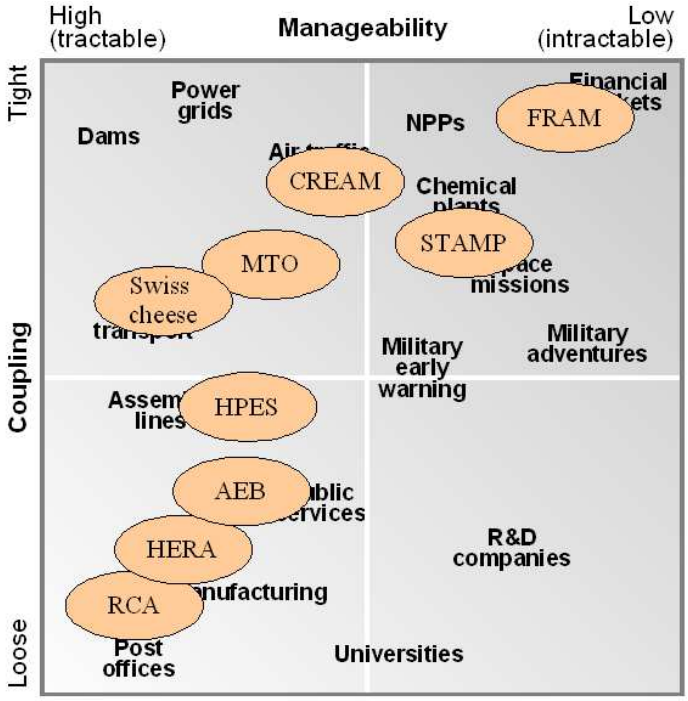

Figure 3: Characterisation of accident investigation methods

It makes sense that any method would be just about adequate for the typical type of problems at the time it was developed. Indeed, there would be little reason to develop a method that was more complex or more powerful than required, not least because it would be difficult to imagine what that should comprise. As argued in the beginning, new methods are usually developed because the existing methods at some point in time encounter problems for which they are inefficient or inadequate. This, in turn, happens because the socio-technical systems where accidents happen continue to develop and to become more complex and more tightly coupled. The inevitable result is that even new methods after a while become underpowered because the nature of the problems change, although they may have been perfectly adequate for the problems they were developed for in the first place.

The position of the various methods on the diagram in Figure 3 presents a characterisation of the methods using the two dimensions of coupling and manageability, and thereby indirectly represents the developments of socio-technical systems since the 1980s – and indeed since the 1930s. Without going into the details of this development, the lower left quadrant can be seen as representing industrial systems before the middle of the 20th Century, i.e., before the large scale application of information technology. The development since then has been one of tighter coupling (moving up into the upper left quadrant) and a loss of tractability (moving right into the upper right quadrant). This has in turn required the development of new methods, as shown in the diagram.

The position of a method reflects the assumptions behind the method, specifically what has been called the accident model. The arguments for each method were presented above. To illustrate the significance of the position, consider for instance the two extremes RCA and FRAM.

- Root cause analysis (RCA) assumes that adverse outcomes can be described as the outcome of a sequence (or sequences) of events or a chain (or chains) of causes and effects. The investigation is therefore a backwards tracing from the accident, trying to find the effective cause(s). The method requires that the system is tractable, since it otherwise would be impossible to carry out this backwards tracing. The method also requires that the system is only loosely coupled, since it otherwise would be impossible to feel confident that the correction or elimination of the root cause would prevent a recurrence of the accident.
- The functional resonance accident model (FRAM) assumes that adverse outcomes are the result of unexpected combinations of normal variability of system functions. In other words, it is the tight couplings that lead to adverse outcomes and not sequences of cause(s) and effect(s). Since the investigation furthermore looks for functions rather than structures, it is less problematic if the description is intractable. Indeed, functions may come and go over time whereas system structures must be more permanent. Functions are associated with the social organisation of work and the demands of a specific situation. Structures are associated with the physical system and equipment, which does not change from situation to situation.

This characterisation does not mean that FRAM is a better method than RCA in an absolute sense. (A similar argument can be made for any other comparison of two methods.) But it does mean that FRAM is well-suited for some kinds of problems and that RCA is well-suited for others, more precisely that FRAM is better suited for risks in tightly coupled, intractable systems. (It of course also means that there are problems for which either method is ill-suited.)

The risks that dominate in present day systems have a different aetiology than the risks that dominated one or two decades ago. This has two important ramifications. The first is that it is more difficult to understand these risks. It is harder to understand that risks may exist, at least until an accident has happened. It is harder to understand the "mechanisms", because risks can arise from non-linear interactions among normal performance variability as well as from consequences of failures and malfunctions. And because of that it is also more difficult to think of ways to reduce or eliminate the risks. In tractable systems, risks are often associated with specific components or subsystems, or with specific actions or operations. Risk reduction can therefore be achieved by either eliminating the risk, by preventing certain actions, or by protecting against the outcomes. But only the last option is available for intractable systems. Eliminating or preventing performance variability may well reduce the risk, but it will also impede normal functioning.

The second ramification is that many of the established risk assessment and accident investigation methods are inadequate for tightly coupled, intractable systems. This dilemma

was made clear when Perrow proposed that accidents could be seen as normal, because risk assessment and accident investigation methods naturally focus on that which is abnormal or dysfunctional. The lesson to be learnt from that is that we must continue to evaluate critically the methods that are at our disposal. The fact that a method has worked in the past is no guarantee that it will also work in the future. The development of new socio-technical systems means that new risks will emerge, and therefore that existing methods sooner or later will need to be complemented with more powerful approaches. What these will be, no one can say for certain.

References

Adamski, A. & Westrum, R. (2003). Requisite imagination. The fine art of anticipating what might go wrong. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (pp. 193-220). Mahwah, NJ: Lawrence Erlbaum Associates.

Arrhenius, S. (1896). On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine and Journal of Science, Series 5, Volume 41*, pp. 237-276.

Bento, J.-P. (1992). *Människa, teknik och organisation. Kurs i MTO-analys för Socialstyrelsen*. Studsvik, Nyköping: Kärnkraftsäkerhet och Utbildnings AB.

CISHC (Chemical Industry and Safety Council), (1977). *A guide to hazard and operability studies*. London: Chemical Industries Association.

Cooper, S. E., Ramey-Smith, A. M., Wreathall, J., Parry, G. W., Bley, D. C. & Luckas, W. J. (1996). *A Technique for Human Error Analysis (ATHEANA)*. Washington, DC: Nuclear Regulatory Commission.

Dekker, S. (2006). *The field guide to understanding human error*. Aldershot, UK: Ashagte.

Fujita, Y. & Hollnagel, E. (2004). Failures without errors: Quantification of context in HRA. *Reliability Engineering and System Safety*, *83*, 145-151.

Heinrich, H. W. (1931). *Industrial accident prevention*: New York: McGraw-Hill.

Helmreich, R. L., Merritt, A. C. & Wilhelm, J. A. (1999). The evolution of Crew Resource Management training in commercial aviation. *International Journal of Aviation Psychology*, *9*(1), 19-32.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method*. Oxford, UK: Elsevier Science Ltd.

Hollnagel, E. (2004). *Barriers and accident prevention*. Aldershot, UK: Ashgate.

Hollnagel, E., Woods, D. D. & Leveson, N. G. (2006). *Resilience engineering: Concepts and precepts*. Aldershot, UK: Ashgate.

Hudson, P., Primrose, M. J. & Edwards, C. (1994). *Implementing tripod-DELTA in a major contractor*. In: Proceedings of the the SPE International Conference on Health, Safety and Environment, Jakarta, Indonesia. Richardson, TX: Society of Petroleum Engineers.

INPO (1989). *Human performance enhancement system: Coordinator manual* (INPO 86-016, Rev. 02). Atlanta, GA: Institute of Nuclear Power Operations

Isaac, A., Shorrock, S. & Kirwan, B. (2002) Human error in European air traffic management: The HERA project. *Reliability Engineering and System Safety*, *75*(2), 257-272.

Le Bot, P., Cara, F. & Bieder, C. (1999). *MERMOS, A second generation HRA method*. Proceedings of PSA '99, International Topical Meeting on Probabilistic Safety Assessment", Washington, DC.

Leveson, N. G. (1995). *Safeware - system safety and computers*. Reading, MA: Addison-Wesley.

Leveson, N. G. (2004). A New Accident Model for Engineering Safer Systems. *Science*, *42*(4), 237-270.

MIL-STD-1629A (1980). *Procedures for performing a failure mode, effects and criticality analysis*. Washington, DC: Department of Defence.

Nouvel, D.; Travadel, S. & Hollnagel, E. (2007). *Introduction of the concept of functional resonance in the analysis of a near-accident in aviation*. Ispra, Italy, November 2007, 33rd ESReDA Seminar: Future challenges of accident investigation.

Perrow, C. (1984). *Normal accidents*: *Living with high risk technologies*. New York: Basic Books, Inc.

Pringle, J. W. S. (1951). On the parallel between learning and evolution. *Behaviour*, *3*, 175-215.

Reason, J. T. (1990). *Human Error*. Cambridge University Press

Reason, J. T. (1997). *Managing the risk of organisational accidents*. Aldershot, UK: Ashgate.

Rollenhagen, C. (1995). *MTO – En Introduktion: Sambandet Människa, Teknik och Organisation*. Lund, Sweden: Studentlitteratur.

Sawaragi, T.; Horiguchi, Y. & Hina, A. (2006). *Safety analysis of systemic accidents triggered by performance deviation*. Bexco, Busan, South Korea, October 18-21. SICE-ICASE International Joint Conference 2006.

Svensson, O. (2001). Accident and Incident Analysis Based on the Accident Evolution and Barrier Function (AEB) Model. *Cognition, Technology & Work*, *3*(1), 42-52.

Takano, K., Sawayanagi, K. & Kabetani, T. (1994). System for analysing and evaluating human-related nuclear power plant incidents. *Journal of Nuclear Science Technology*, *31*, 894-913.

Wears, R. L., Perry, S. J., Anders, S. & Woods, D. D. (2008). Resilience in the emergency department. In Hollnagel, E., Nemeth, C. & Dekker, S. (Eds)., *Remaining sensitive to the possibility of failure*. Aldershot, UK: Ashgate.

Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. (1999). Organising for high reliability: processes of collective mindfulness. *Research in Organisational Behaviour*, *21*, 81–123.

Westrum, R. (2006). A typology of resilience situations. In E. Hollnagel, D. D. Woods & N. G. Leveson (Eds.), *Resilience engineering: Concepts and precepts*. Aldershot, UK: Ashgate.

Wilson, P. et. al., (1993). *Root cause analysis – A tool for total quality management*. Milwaukee, WI: Quality Press.

Yoshizawa,Y. (1999). *Activities for on-site application performed in human factors group*. Proceedings of 3rd International Conference on Human Factors in Nuclear Power Operation (ICNPO-III), Mihama, Japan.