Title	Human and organizational biases affecting the management of safety
Author(s)	Reiman, Teemu; Rollenhagen, Carl
Citation	Reliability Engineering and System Safety vol. 96(2011):10, pp.1263-1274
Date	2011
URL	http://dx.doi.org/10.1016/j.ress.2011.05.010
Rights	Copyright © 2011 Elsevier. Reprinted from Reliability Engineering and System Safety. This article may be downloaded for personal use only

Reliability Engineering and System Safety 96 (2011) 1263–1274

Human and organizational biases affecting the management of safety

Teemu Reiman[a*], and Carl Rollenhagen[b]
[a]VTT, Espoo, Finland
[b]KTH, Stockholm, Sweden

ABSTRACT

Management of safety is always based on underlying models or theories of organization, human behavior and system safety. The aim of the article is to review and describe a set of potential biases in these models and theories. We will outline human and organizational biases that have an effect on the management of safety in four thematic areas: beliefs about human behavior, beliefs about organizations, beliefs about information and safety models. At worst, biases in these areas can lead to an approach where people are treated as isolated and independent actors who make (bad) decisions in a social vacuum and who pose a threat to safety. Such an approach aims at building barriers and constraints to human behavior and neglects the measures aiming at providing prerequisites and organizational conditions for people to work effectively. This reductionist view of safety management can also lead to too drastic a strong separation of so-called human factors from technical issues, undermining the holistic view of system safety. Human behavior needs to be understood in the context of people attempting (together) to make sense of themselves and their environment, and act based on perpetually incomplete information while relying on social conventions, affordances provided by the environment and the available cognitive heuristics. In addition, a move toward a positive view of the human contribution to safety is needed. Systemic safety management requires an increased understanding of various normal organizational phenomena – in this paper discussed from the point of view of biases – coupled with a systemic safety culture that encourages and endorses a holistic view of the workings and challenges of the socio-technical system in question.

Keywords: Safety management, bias, safety science, organizational factors, human factors

[*] teemu.reiman@vtt.fi. Tel: +358 50 3427 268 Fax: +358 20 722 5888

1. Introduction

Management of safety is always based on underlying models or theories of organization, human behavior and system safety. These theories are either explicit or implicit, or a combination of both. An important function of theories and models of safety management is that they create expectations and suggest potential actions. Thus, they direct attention to certain issues and away from other issues, and make certain solutions seem more relevant than others. If they lead to actions that do not contribute to safety or actually create harm, we can label these models biased. The aim of this article is to review and describe a set of potential biases in safety management approaches and their possible consequences for safety. We have tried to extract what safety management professionals and researchers generally look for – and what they might miss.

We will focus on biases that have relevance to safety management in a broad sense. The concepts of a Safety Management System (SMS) and Safety Management (as an activity within a SMS) have various definitions in the literature and no consensus exists about the precise content and scope of these terms [9, 50]. However, a tentative view of these terms suggests that SMS is associated with policies, objectives, procedures, methods, roles and functions that aim at controlling hazards and risks in socio-technical systems. Hale et al. [21, p. 121] have described SMS as "a set of problem-solving activities at different level of abstraction in all phases of the systems life cycle". In this paper we focus on a subset of important activities in SMS systems: experience feedback activities (including event investigations), risk analytic activities, continuous development, safety indicators, organizing and the content of safety policies. We discuss how various biases might influence the content and scope of these SMS activities.

1.1. The significance of beliefs and assumptions in safety management

The validity of the theories underlying safety management activities greatly contributes to the effectiveness of safety management. Major accidents have challenged the general conceptions and presumptions about safe and effective operations. The underlying models of safety were in these cases proved wrong – at least until the accident was explained in hindsight as fitting an existing paradigm. Thus, not even accidents have always been sufficient to prove the safety theories wrong [cf. 65]. Some of the reasons for this have to do with the biases related to human and organizational behavior, such as hindsight bias and attribution error, which are tackled in this paper.

Erik Hollnagel [24] has used the phrase "What–You–Look–For-Is-What-You-Find" to illustrate the effect of priori assumptions and models on the findings of e.g. accident investigations. When people formulate expectations, they assume that certain sequences of actions/events are likely to happen. Such expectations and their associated assumptions are partly embedded in organizational practices, routines, norms and management strategies [74]. Expectations create orderliness and predictability, and offer guidance for performance and interpretation. Expectations guide our attention and search for evidence, thus making it easier to confirm the accuracy of our original expectations by neglecting contradictory information. Expectations can also undermine reliable and resilient performance because they encourage confirmation seeking, reliance on existing categories, and oversimplification. Consequently, organizations should continuously work to override e.g. the typical human tendency (a bias) to seek confirmation and avoid disconfirmation [74].

There is plenty of evidence that the underlying assumptions and theories in use among safety professionals as well as safety researchers vary a lot. For example, Korolija and Lundberg [34] have noted in their study of professional accident investigators that there was no such thing as a professional usage of the concept of "human factor" but a spectrum of meanings among the investigators [see also 35]. Steele and Pariés [62] have studied safety beliefs in the aviation industry. They point out that some of the common assumptions about aviation safety [prevalent in the field] are either false or do not apply under certain conditions. They further argue:

> "Examples of the kind of assumptions we are referring to are: 'humans are a liability (and therefore automating the human out of the system makes it categorically safer)' or 'accidents

> occur as a linear chain of events' or 'following the procedures guarantees safety', etc. Many of the models and methods currently in use are based on these assumptions, and, therefore, they do not meet the needs of the modern aviation industry – they may in fact prevent further progress. … Most worrying of all is the fact that these assumptions are tacit: they are assumed to be 'truths' and are taken for granted without most people even being aware of them or considering them possible points for debate. An example is the notion that 'every accident has a cause'."[62]

A recent interview study in air traffic management (ATM) and airport operations [64] illustrated that managers' conceptions regarding human factors were dominantly individual and error based. However, wider and more systemic conceptions also surfaced during the course of the interviews, but there was large variance between managers in their conceptions. According to the study [64, p. 445], uninformed, individual or error-based conceptions are "insufficient or overly simplified in the context of ATM and airport operations". The study concludes that a "human factors strategy" would be needed in the target organization to form more congruent conceptions among personnel.

Different industries seem to exhibit more or less maturity in thinking about 'human error' as a contributing factor to negative events. For example, in a study about accident investigation practices in various industries in Sweden (e.g. nuclear, transportation, patient safety, etc) it was found [53] that investigators in some branches (e.g. rail) tended to believe on individual error as a cause of events. Investigators in some other branches were more attentive to various contextual factors that influenced human performance. An interesting question in the context of this finding is to what extent a less mature view of human error should be explained by the existence of collective biases in thinking about human performance. In addition, it is well known that too biased a view of human performance contributes to the development of an organizational culture where people are reluctant to report negative events since they are afraid of being exposed to blame as well as to attempts to decrease the role of humans in the production process by e.g. automation. Thus, biases related to human and organizational behavior can be assumed to have wide implications in the design and overall functioning of the entire socio-technical system.

1.2. Thematic areas of safety beliefs and the aims of the study

We have selectively aggregated information from literature on human sciences (psychology, sociology, human factors) together with current safety science literature and our own studies in order to abstract important lessons for safety management. Instead of offering an additional set of definitions to the scientific debate, we aim to illustrate the various issue domains which the ambiguities reflect. Thus, we will focus on the phenomena rather than the concepts that have been used to describe them. Figure 1 illustrates some of the questions that safety professionals need to find an answer to in order to carry out their work. Many of the questions are such that professionals do not explicitly think about them; rather they might have an implicit answer that guides their safety work.

Figure 1. *An illustration of the safety management biases differentiating four interrelated thematic areas*

We will outline human and organizational biases that have an effect on the management of safety in four thematic areas: beliefs about human behavior, beliefs about organizations, beliefs about information and biases in safety models. Each of the thematic areas includes a number of biases that are elaborated below. The extracted biases are based on our experience in various research and development projects in different safety critical environments [see e.g. 35, 38, 46, 47, 48, 49, 51, 52, 53] as well as a review of the relevant literature[1]. We acknowledge that the topic of human and organizational biases in safety management is such a vast one that it is impossible to cover all aspects of it in detail. However, our aim is to provide an overview of the thematic areas where the biases manifest themselves as well as encourage reflection on the effects of these biases on safety management.

The biases are all interrelated and thus the four thematic areas also partly overlap. Underlying conceptions concerning the nature of human behavior have an effect on how the safety practitioner views the organization, and these views in turn affect notions of safety and ways of gathering information on safety. Further, implementing methods and models incorporating certain biased assumptions may slowly affect how the practitioner views human conduct or organizational performance.

2. Beliefs about the nature of human behavior

Despite the growing awareness of the importance of so-called human factors, views about how humans contribute to safety often remain negative among practitioners as well as researchers [cf. 34, 45]. In addition, the practitioners' view of the 'human factor' is very much based on the idea of a single human being responsible for incidents and accidents [34], rather than one where a more social and systemic framework is adopted. In this section we discuss the biases that shape this view.

2.1. Human performance and human error

[1] Preliminary results were presented in a paper 'Identifying the typical biases and their significance in the current safety management approaches' presented at the 10th International Probabilistic Safety Assessment & Management Conference, 7-11 June 2010, Seattle, USA.

It is often argued that over 80 percent of all accidents are caused by human errors or unsafe acts. This statement may seem reasonable at first sight, since humans design, construct, operate and maintain socio-technical systems. This assumption, which dates back almost a hundred years, still constitutes a basis for many safety initiatives [36]. The abundance of human error is often used as a justification for various "softer" methods such as behavioral programs or human factors training. In fact, in the nuclear industry, for example, the entire concept of "human performance" is sometimes understood as basically being error prevention programs and techniques [47]. People are seen as a threat to safety because they may perform unexpected actions. This makes the *reduction of variation in human behavior* one of the main challenges (though often visible only between the lines). It is undisputable that the actions of humans do not always fulfill their intended goals or that sometimes humans clearly err in their decision-making or make slips. These phenomena are commonly associated with the concept of "human error". However, the concept of 'human error' does not explain past incidents or predict the future any better than the term 'technical failure' explains or prevents breakdowns [38]. The propensity to label negative outcomes as due to individual error says more about general human tendencies in attributing causality than about the event itself. The attribution of error is a (social) judgment about human performance made with the benefit of hindsight [75]. A more fruitful starting point would be to treat human performance variability as a normal phenomenon behind both success and failure [24].

Recent research has provided compelling evidence that decision-making in natural work situations is seldom synonymous with conscious selection between different alternatives. The available tools, the environment, people and the terminology used affect the perceptions and interpretations of individuals [28, 33, 38]. Furthermore, risk perception is influenced by the employee's duties, his or her department and work role [1]. Thus, people may observe risks in their organization in systematically different ways. Also, the heuristics described later in this section affect the way personnel perceive and evaluate risks. The same applies to emotions; an agitated person can estimate a risk to be high as likely as a person can estimate a high risk and become agitated because of the high risk [63]. In fact, all action is affected by emotion, for better or worse, depending on the situation.

Human behavior is always contextual. This means that humans act based on affordances provided by their environment (in terms of cues for action embedded in tools and technology) and expectations and norms provided by significant others (the peer group). Human behavior is dictated by the local rationality principle: people (usually) do what makes sense given the situational indications, operational pressures and organizational norms existing at the time [13, p. 12]. We will return to the contextual and locally rational nature of human behavior on a number of occasions in this article.

2.2. Reasoning and inference

Human reasoning is very different from that of a machine (e.g. a computer). People act and make decisions not on the basis of careful and systematic data processing but rather by pattern recognition (see 4.1 below) and the use of certain heuristics. The *confirmation bias*, or the tendency to look for (and find) information that confirms expectations and disregards information that negates them, was already mentioned in the introduction. The *availability bias* refers to the finding that the ease with which instances of the event or issue can be recalled from memory affects the estimation of its frequency in general. For example, the sorts of disasters that receive a lot of media attention are later judged as more frequent than they are in reality due to the vivid images easily produced from the memory. *Anchoring* refers to the human tendency to use reference points and benchmarks, even arbitrary ones if necessary. [16, 30] Next we will look more closely at one of the most fundamental human tendencies, the attribution of reasons to behavior.

People have a general tendency to perceive others as having quite stable traits, and see their behavior as less dependent on context than their own [16]. Thus, the errors and mistakes of others are perceived as being due to stable traits ('bad apples') rather than contextual reasons. Approaches having this bias often emphasize the importance of individuals' own attitudes to safety behavior. This is visible e.g. in some occupational safety campaigns where the explicit message is 'you all have the necessary

competence and know how to act safely, thus it is only a matter of attitudes whether you decide to work safely or not'. The so-called behavior-based safety approaches (BBS) essentially share the same underlying logic [27, 59].

Safety science research has argued for a shift to a no-blame approach to safety [12, 44]. Adopting a 'no-blame culture' is an idealistic approach worth discussing in more depth since the no-blame approach contradicts other goals and beliefs such as that people should be accountable for their actions. The basic issue concerns *under what conditions* a person should be held responsible. This, however, is a much more intriguing and difficult question in comparison with the general quest for non-blame cultures. Reason [44, 45], for example, talks about a 'just culture' to highlight the trade-offs between reasonable and unreasonable blame.

2.3. Attribution of causes and causality

A typical human (and organizational) characteristic is a tendency to blame someone else's failure or error on the basis of character (laziness, indifference, lack of ability), instead of situational or work conditions. However, people have been found to explain and justify their own behavior differently. The fundamental attribution error is a tendency for people to over-emphasize dispositional, or personality-based (internal) explanations for behaviors observed in others while under-emphasizing situational (external) explanations [16]. People have an unjustified tendency to assume that another person's actions depend on what 'kind' of person that person is rather than on the social and environmental forces influencing the person. However, this same tendency does not apply to one's own behavior when that behavior is considered successful [16]. Thus, people claim more responsibility for successes than for failures. Success is attributed to skill, failure to randomness (situational, unpredictable influences). This tendency is good for our self-esteem, but bad for learning from experience (see Section 3.2). This bias also seems to operate on the level of social identities, which means that the successful actions of one's own group are considered to be due to the group's characteristics, whereas failures are attributed to external conditions [47].

In hindsight, after an accident many of the weaknesses that exist in organizations are usually revealed. For example, it is quite common to detect 'deviations' from rules and regulations. Some accident investigations practices equate a deviation with a 'cause': however, the fact that something did deviate from a prescribed rule is not necessarily a contributor to an accident or even an abnormal event. On the contrary, routine noncompliance with written procedures and local adaptations are often the norm rather than the exception [10, 67]. Bourrier [10] has argued, in the context of nuclear power plant maintenance, that "local adjustments to and re-arrangements of, rules and, at times, even rule violations, are not only constant but necessary for organizations to effectively pursue their goals". Only when things go wrong are these adjustments considered negative.

People are also quick to make causal inferences between visible events that take place successively. The first event is judged to cause the second. Studies show that the more serious the situation is (for the individual or society), the less attractive the idea becomes that the situation (e.g. an accident) was due to pure chance [38]. Chance also implies that the same incident could target or could have targeted me. This is why people so readily stress the fact that an incident could have been prevented and the person involved must have caused it [16, 37]. Looking for human errors after an event is a 'safe' choice, since one always finds them in hindsight. Looking and finding human errors makes it easier to find out who is responsible for the accident, who should be held accountable, and where preventative measures should be targeted. Unfortunately, the preventative measures are usually off target if 'the cause' has been attributed to individual error. Accidents occur due to a combination of many factors, which are not necessarily dangerous or erroneous in isolation but when their influences combine, they may expose the organization to an accident. By blaming the individual, people can maintain the assumption (or illusion) that the system is basically safe, or will be as soon as it can get rid of the 'bad apple' [cf. 4, 12]. Thus, identifying where the responsibility for incidents lies among individual decision-makers allows for quick (and 'dirty') remedies such as firing, transferring or retraining them [65]. A problem with these remedies is that they are seldom very effective in achieving the goal of

increased safety and long-term productivity. *Hindsight bias* refers to a finding that it is very difficult for people to ignore the knowledge of an actual outcome to generate unbiased inferences about what could or should have happened [16, p. 193]. Finding out that an outcome has occurred increases its perceived likelihood. People are, however, unaware of the effect that outcome knowledge has on their perceptions [15, p. 310].

3. Beliefs about the nature of organizations

Perspectives on the nature of the organization vary from those that emphasize their social and interpretive aspects to more rationalistic approaches focusing on structures and official routines [22]. Many theories of accidents and safety in industrial organizations are based on a static and rational model of an organization as elaborated in this section.

3.1. Causal primacy of structure

Researchers and practitioners have a tendency to view social phenomena such as safety management in individual terms and in terms of structure (e.g. an error) instead of process (e.g. performance variability). Often, this tendency is due to the fact that individual phenomena and individual behavior are more immediately visible and apparent than social phenomena. In addition, structures are more stable (by definition) than processes and as such they have been easier to study.

Many current models of safety management are based on a rational or a non-contextual image of an organization [46]. They thus originate from a "traditional" mechanistic paradigm of organization science [12, 68]. In this paradigm organizations are considered mechanistic and essentially rational. The underlying assumption is that the purpose and goals of the organization are self-evident and explicit for everyone. Organizational routines are considered well-defined, regular and stable forms of behavior used to accomplish organizational goals. Procedures and instructions are utilized to define appropriate behavior and its outcomes. The role of management and management systems in supervising and directing organizational behavior is emphasized. This rational-instrumental theory of an organization is based on the assumption that people set explicit goals, make rational choices and act on the basis of objective facts [46]. Waring and Glendon [69] criticize safety management systems that are based on an overly rational image of the organization and argue that they may be only partly effective, while creating the illusion that the risks have been fully controlled [see also 12, 68].

The reality of organizational life is usually very different from that described in formal documents. This is natural in all social contexts and not necessarily a bad thing. The search for deviations from the prescribed logic of the organization may in fact camouflage the reality since causes are attributed based on observed deviations rather than exposing contextual factors that unfold the reality of organizational activities. Antonsen et al. [4] point out that "one of the far-reaching consequences of such rationalistic approaches [that view safety as compliance with the official procedures] is that planning and pre-programming are separated from the people performing the work". This can lead to a gap between work as imagined (by management and planners) and work as actually done (on the shop floor). Too strict planning and proceduralization can demotivate the people doing the work [20]. This can lead to employees finding creative ways of 'enriching' their jobs or shortcuts and rule violations to make the work easier or more efficient. These solutions usually remain informal parts of the employee work culture, often unknown to distant (safety) managers. Any endeavor to design the work for error-free performance and standardized outputs actually plants the seed of its own failure by making the work (as imagined) rigid, inflexible and predictable in an environment that is inherently dynamic and unpredictable [20, 24, 61].

3.2. Organizational change and learning

Some approaches to safety management aim at guaranteeing that nothing has changed, and that all the safety measures are still in place. These approaches do not typically acknowledge the inherent change of sociotechnical systems and the fact that yesterday's measures may be today's countermeasures [47].

Thus, they are also often based on under-specified safety models (see Section 5.1) in addition to having a static view on organization. These approaches are based on the notion that organizations only change when the management decides upon a new structure or process. Otherwise the organization is statically carrying out its tasks in the way the processes dictate, or, at least, that is the implication.

Weick [70, 71] has emphasized that instead of speaking of *organization*, we should speak of *organizing*. What we perceive as an organization is the (temporary) outcome of an interactive sense-making process [70]. Even heavily procedural socio-technical systems adapt and change their practices locally and continually [cf. 10, 12]. Routines and practices develop over time even without any noticeable pressure on change. People optimize their work practices, come up with shortcuts to make their work easier and more interesting, lose interest in commonplace and recurrent phenomena, and have to make tradeoffs between efficiency and thoroughness in daily tasks [24]. Sometimes change is needed in order to keep things stable in the organization, i.e. to counteract external change pressures and the internal gradual drift of work practices. If everyday work requires too much adaptation and improvisation then the system is unstable and change management is needed to stabilize it. Snook [61, p. 194] has defined "practical drift" as the slow steady uncoupling of practice from written procedure, in which, after extended periods of time, locally practical actions within subgroups gradually drift away from originally established procedures [cf. 13]. According to Snook, constant demands for local efficiency dictate the path of the drift. Thus, drift in an organization is about constant organizing [cf. 70] in face of everyday challenges – learning to adapt and adapting to be able to carry on.

Another bias related to learning is connected to the idea that organizations (and humans) learn only from failures and mistakes, and not from the daily successes and "non-events". This view also typically does not question the definition of what constitutes a failure in the first place. However, defining failure and success are social and political processes, and by reinterpreting history, each can be turned into the other [65, 71]. Sagan [54] reminds researchers and practitioners of "the resourcefulness with which committed individuals and organizations can turn the experience of failure into the memory of success". For successful organizations the danger lies in developing a complacent attitude if the future is considered an automatic repetition of history. Learning is more than an accumulation of knowledge; it involves continuous change and the development of thinking (and action) in a specific operating environment. Nor does learning simply mean accumulation of (work) experience. Long experience does not necessarily and automatically lead to more advanced models of thinking and action, but may rather result in restricted routines that are difficult to change. Learning is also dependent on one's view of information and orientation towards uncertainty [37], see Section 4.

3.3. Neglect of emergent phenomena

The phenomenon of emergence refers to patterns, structures or properties emerging at the system level (e.g., an organization) that are difficult or impossible to explain in terms of the system's components and their interactions [55]. Emergent phenomena cannot be reduced to the properties or functioning of its components. For example, on an individual level, mental properties may not be easily reduced to neurobiological processes. On the organizational level emergent phenomena include shared beliefs and practices (culture) as well as work climate [55]. These emergent phenomena affect performance at the individual level in a process called downward causation. Safety can also be considered an emergent phenomenon, making a systems view an imperative if the aim is to evaluate or develop the safety of the entire sociotechnical system (system safety). This means that safety cannot be understood or managed by understanding or managing its constituent parts in isolation. As safety is a property of the sociotechnical system, all attempts to understand or control systems by reducing them to individual components are subject to the bias of neglecting emergence. According to this bias, "the functioning or non-functioning of the whole [complex sociotechnical system] can be understood through the functioning or non-functioning of [its] constituent components" [13, p. 73].

Organizational culture is a term that has often been used to denote the emergent phenomena in workplace settings. For example, organizational culture as an emergent phenomenon can be proposed

to have an influence on anyone working in the organization –influence that is either positive or negative in terms of safety outcomes. Weick has pointed out that "strong cultures can compromise safety if they provide strong social order that encourages the compounding of small failures" [72, cf. 54] and further that "organizations are defined by what they ignore – ignorance that is embodied in assumptions – and by the extent to which people in them neglect the same kinds of considerations" [72]. Dekker [13] also views *drift* (see above, Section 3.2) as an emergent property of a system's adaptive capacity – drift is a normal by-product of organizational activity. In a certain sense, all organizational practice can be considered emergent. Practice is something that results from the interaction of people with their environment and its regularities and this interaction cannot be fully understood without considering the practice as something that has been developing gradually in the organization and something that simultaneously guides individual action and is produced by individuals acting.

A climate of competition and acute cost awareness affects the way companies conduct business, but few accident investigations would blame such broad system factors (e.g. capitalism) when attributing causes for accident [cf. 26]. However, Rasmussen and Svedung [43] have for example argued that economic factors such as cost pressures in competitive environments are significant contributors to large-scale accidents. They do not, however, take any political stance: rather they assume the economic environment as given and discuss the consequences for safety management in a "dynamic society". The "political" dimension is often lacking in much safety science research and discussion [cf. 3].

4. Beliefs about the nature of information and uncertainty

Zio [78, p. 136] argues that "in spite of how much dedicated effort is put into improving the understanding of systems, components and processes through the collection of representative data, the appropriate characterization, representation, propagation and interpretation of uncertainty will remain a fundamental element of the reliability analysis of any complex system." Uncertainty is often categorized into two distinct types. The first type of uncertainty is called epistemic uncertainty and it refers to lack of knowledge or information on the object of study. The second type is called aleatory or stochastic uncertainty and it refers to randomness due to inherent variability in the system. This type of uncertainty cannot be reduced by further data gathering and information acquisition. [5, 78]

4.1. Over-quantification

Being able to quantify a variable is often perceived as a characteristic of control. The validity and reliability of various measures associated with risk and safety (and their associated models) is a much-discussed subject in safety science. Naturally, numbers are no problem (or a solution) in themselves but rather how one makes sense of these numbers (including how the numbers have originally been obtained). In particular it is important, as far as possible, to make hidden assumptions explicit and to reveal the uncertainties associated with safety measures. For example, safety culture surveys might have a response bias due to genuine but incorrect beliefs or an effort to make one look good (impression management). Relying on the overall scores without understanding the uncertainty behind them might be counter-beneficial (e.g. create a false sense of comfort).

Performance indicators that are easy to quantify may also divert attention away from more subtle but important issues - such as issues of power in organizations. Furthermore, many issues of subjective risk and "gut feelings" are difficult to monitor with quantitative indicators. People may experience that 'something is wrong' or 'missing' but not be able to clearly communicate what it is. If an organization consequently dismisses such reports as representing little more than general complaints rather than something that actually could be a vague perception of an existing risk, then subtle but existing risks might prevail although they in fact have already been detected. What we often call 'intuition' is not some mystical faculty of the mind but rather a consequence of experience-based pattern matching that may express itself rather vaguely as a feeling of recognition [32]. Finally, without any underlying model describing the postulated *causal relations* among a set of performance measures it is indeed

difficult to know why a change has occurred (learning) and what the change implies for safety management strategies or safety levels in the organization. Wilkin [74, p. 238] reminds us that "mathematics is an acausal and logical system that can be used for a variety of purposes, including useful ones in the social sciences, but it is severely limited in its uses and says nothing about the crucial question of causality".

Over-quantification also manifests itself in subjects other than performance indicators. One example of over-quantification relates to the "iceberg" folk models of safety (or accidents) where small incidents are counted in the hope of predicting when a more serious event will take place. Another bias involves a type of binary thinking where an existence (1 or 0) of a system or procedure in the organization takes precedence over its actual functioning [cf. 4], e.g. in auditing or development. Subjective and social phenomena are hard to quantify and thus overreliance on quantification may lead to dismissing social phenomena as non-real. Weick [71] reminds us that people's subjective expectations are real in the sense that they have an effect on perception and behavior. The influence of subjective feelings, social norms and climate on organizational safety should not be dismissed, even if the phenomena themselves or their effects cannot be quantified [cf. 65].

Wilkin [74, p. 236] states that critical social scientists (following Bhaskar) call the 'epistemic fallacy' a philosophy (positivism) that espouses the idea that what exists is limited to what can be experienced, ultimately that which can be observed (and counted, we would add). He goes on to state that in this positivistic view "the relationship between the necessary (structural relations) and the contingent (the acts of agents as groups or individuals) disappears". Thus, he argues, positivists can neither make sense of these structural/systemic properties in ontological terms as real things (what kind of things they are), nor can they make sense of them in epistemological terms (how can one gain knowledge about them). Quantification is thus carried out by giving structure causal primacy (Section 3.1) and by neglecting emergent phenomena (Section 3.3).

4.2. Uncertainty, randomness and probability

In the name of safety philosophy, many safety critical organizations argue that it is not acceptable to carry out work if its consequences are uncertain. The premise is that one should never experiment or guess; that when in doubt, ask a person who is better informed [38]. This makes the handling of uncertainties a personal question linked to professional competence. It is important for personnel to understand that uncertainty is never caused by an individual alone but is rather related to the object of the work, such as the condition of the technical systems at a process plant or the reliability of the measurement data in process control. The object of the work contains within in it the notion of uncertainty; the progress and effects of the work can never be fully predicted. This is why employees will probably always experience a certain degree of uncertainty when they are at work. Recognizing and coping with uncertainty is related to the development of expertise [37] and decision-making in general.

Like uncertainty, the concept of probability has many uses and interpretations. Often the probabilities that are used in analyses are based on strong assumptions about the phenomena being studied. For example, the "probabilities" of real-life events, such as human errors or pipe fractures, are estimates of true underlying probabilities [5]. We would need an infinite number of these real-life events to converge the estimator and the real probability of the event. In the case of unique events this type of "relative frequency" perspective does not apply. Aven [5] argues that in these subjective knowledge-based situations "probability is a measure of uncertainty about events and outcomes (consequences), seen through the eyes of the assessor and based on the available knowledge". Thus, the probability assignment is influenced by the assumptions and suppositions of the assessor. In the safety management literature this type of measure is often called "expert judgment". Aven [Ibid.] argues that the concept of risk should not be restricted to an assignment of probabilities: risk assessment should take into account the uncertainties that the probabilities are based on. In fact risk assessments can hide the underlying uncertainties if they are based on assigned probabilities alone [5]. It is important to distinguish uncertainty from probability: uncertainty is connected to the underlying phenomena

whereas probability is a tool to express this uncertainty [5]. Risk in this framework consists of three components, A, C, and U, where U is the uncertainty about A and C (will A occur and what will the consequences C be?), including uncertainty about underlying factors influencing A and B [6]. Thus, risk refers to uncertainty about and the severity of consequences (or outcomes) of an activity with respect to something that humans value [6, 7]. In fact, Aven [5] makes an argument that the aim of risk analysis should be to describe uncertainties regarding observable quantities, that is, phenomena and processes in the real world. Thus, uncertainties reside in the world. Aven [5] reminds us that it is not possible to perform a risk analysis without making assumptions. The extent to which these assumptions are made explicit is critical for the success of risk management.

Grote [17] discusses two basic approaches to managing uncertainty in organizations: (1) minimizing uncertainty, and (2) coping with uncertainty. The first strategy is mainly based on a feed-forward control i.e. relying on planning and the monitoring of plans. By and large, this strategy has been implemented by means of detailed rules and regulations intended to guide actors through complex task domains – little freedom is allowed for local initiatives. The second approach, associated with open system theories, focuses on giving actors freedom to cope locally with uncertainty by means of feedback control. One of the assumptions here is that "local actors need to be given as many degrees of freedom as possible, achieving concerted action mainly through lateral, task-induced coordination" [17]. Another way to frame these two strategies is in terms of the distribution of autonomy and control [18], interpreted as self-determination regarding rules and rules to follow (autonomy) in contrast with control "as the influence on a given situation allowing to reach goals which have been determined either autonomously or by others" [18]. Finding a proper balance among different safety management strategies associated with uncertainty is at the core of many safety problems: a system should be designed to allow for prediction and control but at the same time be flexible enough to adapt, innovate and learn [41, 76]. One of our interests in this line of thinking and one which is of relevance for 'biases' in safety management thinking concerns various beliefs about 'the human factor': if humans are dominantly portrayed as a risk factor, then it makes sense to limit variability to combat 'human error'. But since human initiative and flexibility also provide the fuel for avoiding disasters, a safety management system must provide sufficient flexibility to support safe adaptation and learning.

Sometimes there is an emphasis in safety management on a need for everything to be proceduralized. Correct performance and the correct reaction to any situation can then be defined as strict adherence to the corresponding rules and procedures. This same bias can apply to safety management tools and methods. For example, human error prevention and human performance enhancement tools [cf. 27] often entrench a narrow view of safety and human behavior – it is merely compliance with rules and visible safe behavior (e.g. wearing a helmet). At their best, these tools can improve understanding and play a role in competence development if they are based on a systemic view of safety and used in the right way. Unfortunately, the underlying models of safety are not always made explicit. This 'bounded method focus' hides the uncertainties in the methods and might create a false sense of certainty. This focus also easily confuses the nature of explanation and prediction in sociotechnical systems.

4.3. Confusing measurement, explanation and prediction

Ideas of quantification, linear causality (see Section 5.1), and the mechanistic view of humans and organization share an underlying assumption about the nature of explanation. In this mechanistic view, to explain something means to find the cause for the effect (the phenomenon to be explained). This cause can be isolated and its influence analyzed by reductionism [cf. 13], i.e. by inspecting the system component by component to see each part's effect on the whole. Furthermore, this worldview assumes that the bigger the cause the bigger the effect and vice versa - serious effects (e.g. accidents) are the result of serious causes (e.g. major negligence or ineptitude) [13, 40]. There is also a notion that one can clearly separate analysis (measurement) of a system from an intervention or change in a system. However, both depend on what categories are used and how data is gathered from the system. Furthermore, when social phenomena are measured, the mere act of measurement changes the phenomena. Wilkin [74, p. 238] points out that measurement is an important part of any science, yet it is not the same as explanation. In safety science (including ergonomics to which Wilkin explicitly

refers to) an explanation has to build an expansive, rich, contextualized picture of the event or phenomena concerned; one that can set out the variety of causal mechanisms and processes that over time and space generated the event or phenomena and that can engage with the intentions and meanings of the actors involved [74, p. 238]. When dealing with human interaction in sociotechnical systems the quest for explanations has to be supplemented with the quest for *understanding*. This quest should seek to identify the subjective meanings of personnel, all the while acknowledging that this act of identification is itself interpretation guided by one's own meanings and conceptions [57]. The generalizability of knowledge is not the only criterion of validity. On the contrary, each sociotechnical system is unique and this sets constraints on what knowledge can be generalized outside the context of the given system.

The problem of prediction has plagued social sciences for decades. There have been attempts to make social sciences closer to natural sciences by striving toward more deterministic models which would help in not only explaining but also predicting the behavior of the system. A challenge when working with safety critical systems is that negative predictions should not be validated by empirical evidence – if predictions about future accidents are made, it would be unethical to do nothing about them. In fact, when aiming at safety improvement, "self-invalidating predictions" of future incidents – in the event that current culture does not change – could be considered a tool in development work [46, p. 762]. Problems of prediction notwithstanding, when safety-related problems are identified they have to be 'solved' or coped with in some way or other. Results from the problem-finding processes should thus be transformed into reliable and robust solutions, since there is of course no point in having an effective problem identification process if its output is not used as a basis for remedial actions. Also, it is not uncommon for organizations to collect masses of data but with no or little subsequent utilization. Turning data into information and action often presents a bigger problem than implied in safety management manuals. It is sometimes tacitly assumed that solutions are found more or less directly from the results of the problem analysis. However, it is often the case that, for example, accident investigation manuals give very little attention to problem-solving activities [35, cf. 29] – standard recommendations such as more instructions and more training are not necessarily those that in the long run produce reliable solutions. When issues have become an integral part of a system, removing the original 'causes' does not necessarily remove the issues they have created. Solutions should always be considered at system level, not at the level of individual components – such as incompetent individuals or inadequate supervision. Another bias associated with problem analysis is the emphasis on having 'all the facts' before acting. This is related to an organization's way of dealing with uncertainty. This emphasis can paralyze an organization when it has to deal with issues where it is impossible to ever completely remove the uncertainty by collecting more data. Finally, the selection of countermeasures or corrective actions that match the level of perceived or analyzed problem or threat is a challenging task. There is always the danger of over-reacting or under-reacting to the signals.

5. Biases in underlying safety models

Research has shown that concepts such as human factors, safety management, accident, or safety culture have different meanings, definitions and usages among practitioners as well as within the research community [12, 31, 35, 36, 60, 66]. In this section we will deal with biases in safety models.

5.1. Linear causality

The reciprocal causality of technology and the human elements of the system tend to be neglected in the safety management field. People create technology, structures, and processes, which in turn influence how people think, feel, and act. The members of an organization assign meanings and beliefs to organizational elements (structures, systems and tools, others' behavior) and these assigned meanings in turn influence the ways in which the members behave [2, 59, 68]. Even the technological solutions and tools are given meanings by their designers and users, which affect their subsequent utilization.

As mentioned in Section 1.1, Steele and Pariés [62] discovered in their study that people within the aviation industry held implicit and often unfounded assumptions about safety and accidents. This applies most probably to any other safety critical domain as well. The accident models that personnel have at their disposal incorporate beliefs about accidents and the human contribution to safety. When these beliefs are implicit, they might be "dangerous" in the sense that people might have misconceptions about safety. For example, people may believe that there is a direct one-to-one relation between the elimination of a "cause" and the elimination of a resulting "effect". For instance, models suggesting the causal influence of management decisions on work conditions, producing human errors that lead to accidents, largely *neglect the reciprocal influence patterns* among objects and events both within and between different levels of explanation. This, in turn, may evoke false beliefs about the strengths of the remedial actions suggested as well as about the nature of the identified 'causes'. For example, simplified 'human factor' solutions in terms of 'more training and more instructions' frequently appear in event analysis reports - reports that otherwise might have strong technological biases in terms of identified causes. A large emphasis on redundancy as a safety mechanism can also be traced to linear thinking. Accompanied by a reductionist view of an organization (see Section 3.3.), this can lead to overemphasis on adding barriers to "stop progressions to failure" and redundant components to substitute for broken components [13, p. 63]. The fact that the interaction of the added components with the existing system create more complexity and contribute to the system's opaqueness often remains neglected [13, 40, 54].

Rollenhagen [52] points out that when dealing with people, technology and organizations, we are dealing with causally interdependent categories and it makes little sense to attribute (generic) causal primacy to any of the categories in safety models: "To depart from technology in itself without recognition of its interaction with human and organization makes little sense, and departing from 'culture' in itself without understanding how technology and organizations shape beliefs, moral, values, attitudes and behaviors is also problematic" [52]. Still, as noted by e.g. Hollnagel [23, 25] many accident models share a linear view on causality in describing accident sequences and fail to focus on the dynamic interplay among factors. Root cause analyses often share a linear view of the progression of an accident from one or several 'root' causes to their effect, i.e. the accident itself [cf. 13, p. 65]. Models of linear causality also suggest that the effect is proportional to cause; the bigger the cause the bigger the effect. This means that serious effects (e.g. accidents) are believed to be due to serious, or major, causes (see Section 4.3). Instead, in non-linear systems, according to systemic models, small causes can have arbitrarily big effects [cf. 40]. The outputs are not necessarily proportional to the inputs. Actions that seem locally rational can have globally catastrophic results [13] (see also Section 2.1.).

Causal explanations of incidents and accidents have implications for organizational control [39, 66]. Discovering responsibility for incidents among individual decision-makers prompts remedies such as firing, transferring or retraining the individuals concerned (see Section 2.3.). These methods of 'developing' organizational safety mask important systemic issues behind the incidents. An implicit belief in linear causality is also a common folk model. The same problem occurs in 'deeper' theories of organizational culture [56], where future behavior is predicted as a repetition of past behavior or the manifestation of assumptions born out of past behavior.

5.2. Underspecification of safety models

Barley and Kunda argue that, since the dawn of systems theory at the end of the sixties, "work has slipped increasingly into the background as organizational theory converged on the study of strategies, structures, and environments as its central and defining interests" [8]. Rasmussen also points out that even in the safety critical area, "management theories tend to be independent of the substance matter context of a given organization" [42]. There often seems to be an assumption embedded in the models that 'one size fits all', or in other words, that the models are applicable in any domain or any work situation.

According to many academic organizational researchers, the concept of safety culture has become a catch-all concept for psychological and human factor issues in sociotechnical systems [11, 19, 46]. The concern is expressed that safety culture is not seen as a contextual phenomenon, but as a kind of general ideal model without adequate consideration of the work itself being carried out in the organization in question. Furthermore, the specific features of a safety culture inherent in a specific industry or task may be neglected in general safety management practices. But there is also a problem from the opposite perspective: various arenas of safety have produced more or less self-contained regulatory regimes. Depending on national laws and regulations, we can find that "occupational safety", "patient safety", "radiological safety" are in fact becoming so context-dependent that the common features among these different safeties is getting to be a problem. Various safeties may in fact "compete" with each other in the same organization – attempts to satisfy the demands in one area can lead to sacrifices in another area of safety. A more contextual approach is needed that emphasizes simultaneously the productivity, safety and health of the sociotechnical system, i.e. takes into account the core task of the organization [46]. The organizational core task denotes the objective, constraints and requirements of work in a particular context and its understanding is important for safety managers as well as other personnel. This apparent paradox between context- free generalizations (e.g. general accident models, generic safety culture dimensions etc) and specific context-oriented regulatory demands is by no means easy to address. The safety management systems must be sufficiently fine-grained to incorporate specific hazards, tasks etc. found in various task domains but at the same time general enough to integrate various types of safeties found in many complex socio-technical systems.

Experience often narrows one's point of view to focus on some 'pet' theories or solutions that have worked well in the past. As experience accumulates, people learn what works well and what does not. These 'viable' solutions became personal preferences that are then applied to a wide range of situations. Relying on experience makes sense in many cases, but it has its drawbacks. Experience and the implicit beliefs formed from it have a strong influence on what the safety specialist subsequently pays attention to, what he considers important, and what he ultimately finds out. Dekker and Hollnagel [14] say this: "The greatest risk of folk models is that they appear to make sense, even though statements and conclusions may not be falsifiable. They may therefore seem more plausible than articulated models, since the latter require an understanding of the underlying mechanisms." Dekker and Hollnagel [14] argue that folk models are especially vulnerable to their being overgeneralized and extended to situations that they were never meant to apply to. This is due to their common sense appeal, lack of precision and the inability to disprove them.

It is often implicitly assumed that context-free cures for analyzed events in accident analysis manuals succeed by supplying 'boxes' with texts with remarks such as 'suggestion for remedial action' or 'follow up', representing a context-free chain of activities. The interface problems between the different stages in this process are frequently not mentioned at all, or are merely touched upon. Moreover, the typical manual gives very little advice to the practitioner on how the identified weaknesses can be overcome and how the remedial actions should be selected. Non-contextual models serve the purpose of abstracting common themes among industries but they can be dangerous if they camouflage specific hazards or tasks. Another example of over-generalizing concerns different types of hazards; i.e., an increase in occupational accidents is postulated to mean that the risk of a serious process or production-related accident has increased. These iceberg or pyramid models of safety postulate a causal relation between the causes of small injuries and those of major accidents – a relation that scientific evidence does not corroborate [59].

5.3. The definition and content of safety

It is surprising how often the definition of "safety" is taken for granted [48]. In practice different definitions of safety that are used explicitly or implicitly affect safety management priorities and practices. Many implicit models include the idea of safety as an absence of something or the lack of deficiencies, e.g., the fewer the number of unplanned scrams or INES rated events, the higher the safety level at a nuclear power plant. Another bad example would be to use the number of human

errors propose the safety level. But it might be argued that safety should also refer to the presence of something, and not just absence [23, 52]. Mistaking safety for absence easily leads to taking past success as a guarantee of future safety. The problem with past data is that it can only be used to reject hypotheses (such as "our organization is safe"), not confirm them [63]. This *problem of induction* is closely connected to misunderstanding uncertainty and randomness (see Section 4.3).

Many safety management approaches seem to depart from the often implicit assumption that achieving safety is synonymous with avoiding errors. People are perceived as a threat to safety because they may perform unexpected actions, or neglect or bend the rules, forget things, miscalculate, act before thinking things through, and so on. This makes the reduction in variation in human performance one of the main challenges and goals of management. This is a problematic viewpoint. The variation, adaptability and innovation inherent in human activities enable complex organizations to carry out their tasks. More often than causing hazards, people probably carry out their duties exactly as they should, fixing defects in technology, compensating for bad tool or work design or stopping a dangerous chain of events based on intuition [24, 38]. Human flexibility and ability to adapt to changing circumstances is more often a source of success than failure [24].

A more refined approach to safety in complex systems would be to treat safety as an emergent property of the functioning of the entire sociotechnical system [see also Section 3.3]. If safety is understood as something more than the absence of risk and adverse factors, the methods and models of safety management should also be able to focus on the positive side of safety - on the presence of something [23, 52]. Viewed in this way, the management of safety would focus on increasing the potential for the organization to cope with the work on a daily basis, in addition to constraining unwanted variability with safety barriers and redundancy. This requires a systems view of safety management.

6. Toward a systems view of safety management

With reference to an interview study, Zimmermann et al. [77, pp. 270-271] point out that in the field of aviation "practitioners (i.e., non-experts in human factors and ergonomics, etc.) may not have a coherent, consistent, complete framework guiding how they view and understand safety. They may call up individual ideas from different paradigms or frameworks depending on the situation or the cognitive availability of the idea. There are many possible explanations for this, among them that practitioners may not have or need a coherent framework and may not even be aware when they express contradictory ideas. Or they may realise that there are frameworks but may apply different ones to different situations." This is an important point. We argue that a means forward in controlling the effect of biases is acknowledging that there are different and contradictory ways of viewing situations – and then having the flexibility to move from one interpretation to the other. Biases then become a way to see and accept different points of view on the same topic – and cease to be biases anymore. In understanding today's complex systems, this ability to see contradictory viewpoints and not restrict oneself to the linear and obvious is more important than ever.

Figure 2 summarizes the four thematic areas and the twelve main biases associated with them. As the diagram suggests, the biases are closely related and they influence how the organization carries out its safety management activities. Concepts such as safety, reliability, or human factors are not absolute; rather, organizations construct their meaning and act in accordance with this constructed meaning [46]. For example, if the organization socially constructs a view that the essence of safety is to prevent individuals – considered the weakest links in the system - from committing errors, the countermeasures are likely to be targeted at individuals and include training, demotion and blame. When the view is not based on scientific or empirical evidence and its function is not reflected in the organization we can speak of it as being a bias. These biased beliefs can gradually take root in the culture of the organization and become ways of acting, thinking and feeling [56] in relation to safety issues that are taken for granted. Humans act in context and organizational culture denotes that context in work settings. It is imperative that culture does not constrain human behavior by enforcing concepts

and norms of conduct that are too narrow. The culture should provide a common set of decision principles and shared norms and guidelines, yet leave room for adaptation and individual initiative.

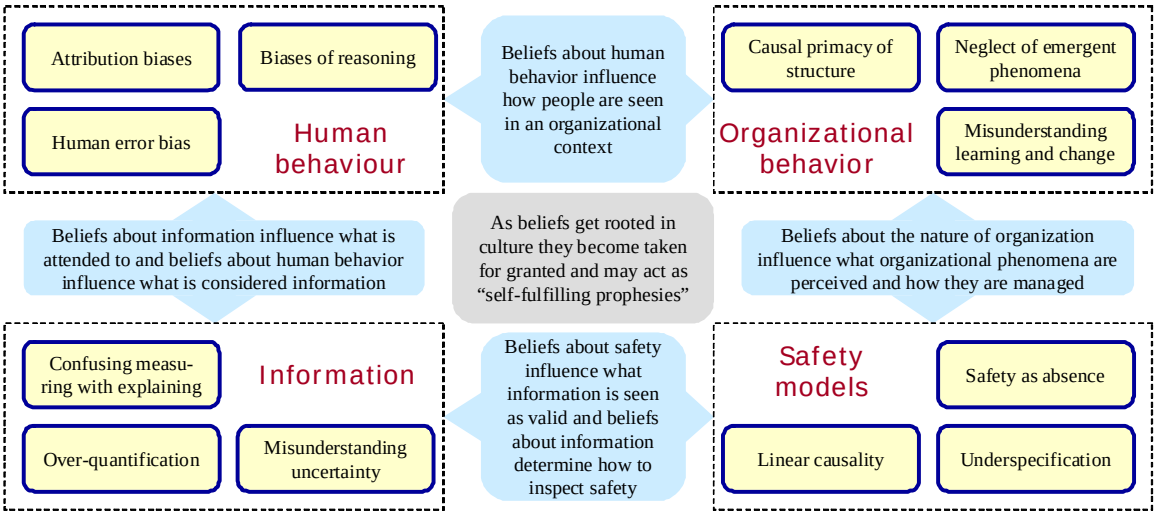

Figure 2. *Summary of the four thematic areas, their associated biases, and interaction*

At worst, the biases depicted in this article can lead to an approach where people are treated as isolated and independent actors who make (bad) decisions in a social vacuum and who pose a threat to safety. Such an approach aims at building barriers and constraints to human performance and neglects the measures aiming at providing prerequisites and organizational conditions for people to work effectively. This reductionist view of safety management can also lead to a separation of so-called human factors from technical issues, undermining the holistic view of system safety.

The four thematic areas and their associated biases have a major influence on how safety management in an organization is defined, executed, and reflected upon. In Figure 3 we have selected some typical elements of safety management [cf. 9, 21, 51, 68, 69] and illustrated how the selected biases from the four thematic areas might have an influence on each of the elements. It has to be remembered that the biases should not in fact be considered in isolation from each other – the biases are interrelated, as are the systems they address. Nevertheless, Figure 3 can be considered as a pragmatic simplification, illustrating how a bias in some area of safety management can have practical consequences for the way safety is managed in the organization.

SAFETY MANAGEMENT	Human behaviour	Organizational behavior	Information and uncertainty	Safety models
Safety policy	**BIAS OF REASONING**: Expressed safety policies focus on most acute things since these are most salient, with "chronic" issues under-emphasized.	**CAUSAL PRIMACY OF STRUCTURE**: Overreliance on technological means for controlling organizational performance	**UNCERTAINTY**: The policy may overemphasize the need for confidence and certainty in an environment where total certainty is impossible	**SAFETY AS ABSENCE**: The policy may overemphasize error prevention instead of safety promotion
Organizing	**HUMAN ERROR BIAS**: The starting point of organizing is in preventing errors and not in creating the context for safety performance	**CAUSAL PRIMACY OF STRUCTURE**: Overreliance on formal organizational structure in controlling safety	**UNCERTAINTY**: Uncertainty management may strive for reduction of uncertainty by standardization and formalization without perceiving the need for flexibility and adaptability.	**SAFETY AS ABSENCE**: Relying on removing the negative may not reveal the underlying dynamics of the system that are essential for creating safety
Risk analysis	**HUMAN ERROR BIAS**: Human influence on safety is considered in negative terms only as potential errors that threaten the otherwise safe system.	**CAUSAL PRIMACY OF STRUCTURE**: Organizational phenomena such as norms and practices are not considered when evaluating the safety of the organization	**UNCERTAINTY**: The estimated probabilities become treated like objective attributes of the environment due to misunderstanding of uncertainty	**SAFETY AS ABSENCE**: Risk analyses may treat the absence of failures as a proof of safety
Experience feedback	**ATTRIBUTION ERROR**: Event investigations seek to identify who made what mistake and attribute blame, instead of identifying how the actors made sense of the situation back then	**NEGLECT OF EMERGENCE**: social phenomena, such as norms and climate, that can contribute to both similar and different events in future are ignored in analyses	**OVER-QUANTIFICATION**: Only quantitative data is considered as valid for risk analysis, neglecting harder to measure human and organizational issues	**LINEAR CAUSALITY**: can lead to focus on either technical, human or organizational factors as having caused the event under investigation
Continuous development	**ATTRIBUTION ERROR**: Development activities might be targeted at changing individual behavior instead of the wider organization due to this bias	**LEARNING AND CHANGE**: A static view on organization ignores the gradual change and optimizing of practices taking place continuously in the organization.	**UNCERTAINTY**: A quest for having all the "facts" before making decisions or improvements may lead to an organizational inability to take action.	**UNDER-SPECIFICATION**: Lack of specification on what type of safety (e.g. process or personnel) an improvement focuses on can lead to wrong conclusions
Safety indicators	**BIAS OF REASONING**: Indicator data can be used to justify preconceptions and expectations that the analyst had before seeing the data, due to a confirmation bias.	**CAUSAL PRIMACY OF STRUCTURE**: Indicators focus on structural elements such as technical reliability, management system, or instructions, neglecting e.g. social issues.	**OVER-QUANTIFICATION**: Only numerical indicators are considered as valid, missing other information such as intuitive judgments, social norms and subjective worries.	**SAFETY AS ABSENCE**: Indicators may be selected to measure negative outcomes (absence) instead of proactive safety activities (presence)

Figure 3: Examples of the consequences for safety of the biases in the four thematic areas concerning the different elements of safety management

An approach that focuses narrowly on 'human error' isolated from the context in which human cognition and behavior occur tends to divert attention away from structural and cultural issues affecting safety. Focus on solely individual issues in continuous development, for example, (Figure 3) can lead to a demand for "good people working harder" in addition to getting rid of the "bad people". This runs contrary to the notion that most errors and accidents are system-induced, and individual compensation for deficient system level solutions will only work up to a point. Often the available methods (and their implicit assumptions) dictate what to look for and analyze, instead of the phenomena dictating what kinds of methods one should utilize. Still, the methods for safety management, including risk analysis techniques and accident investigation tools, typically develop more slowly than the sociotechnical systems they are supposed to help manage. Thus, methods often "lag behind reality" [25]. For example, even though the complexity of modern socio-technical systems has increased, the methods and models used in experience feedback and safety indicators (see Figure 3) are still based on a linear causal view of safety and organizations.

Systemic safety management takes into account people, technology and organization and their interaction on equal terms. More attention needs to be devoted to understanding why things are usually done well enough in the organization instead of only looking at why things went wrong [13, 24]. In addition, in accident investigations it is important to examine why people acted the way they did, and why that made sense to them at the time [cf. 12, 24, 25]. Systemic safety management requires an increased understanding of various normal organizational phenomena – here discussed from the point of view of biases – coupled with a *systemic safety culture* that encourages and endorses a holistic view on the workings and challenges of the sociotechnical system in question.

7. Conclusions

What are the practical implications of our study? Can raised awareness about the phenomena described in this article contribute to better safety management practices? We strongly believe so. The article is surely in itself biased owing to the background and experience of its authors, and by no means claims to be comprehensive or exhaustive in its depiction of biases. Still, we believe that awareness about the phenomena (biases) described in this article and their possible effects may support a more realistic and balanced approach to safety management. For example, an event investigator might check to see whether his conclusions are affected too much by a tendency towards fundamental attribution error, or if what is perceived as a 'deviation' actually is normal variability. A designer of safety performance indicators may, through insights concerning the 'magic of numbers', be more conscious of the notion that 'exact' quantitative numbers might be perceived as being more precise than the underlying phenomena really are. Designers of procedures and safety policies might become better able to design usable products in view of what we know about human behavior. At the same time an understanding of typical human performance biases gives the safety practitioner a better insight into human behavior in complex sociotechnical systems. Human behavior needs to be understood in the context of people attempting (together) to make sense of themselves and their environment, and act based on perpetually incomplete information while relying on social conventions, affordances provided by the environment and the available cognitive heuristics.

It should be noted that relying on the biases described in this article does not always lead to disaster. On the contrary, the strength of the biases lies in the fact that they work *most of the time*. It makes sense to think linearly or extrapolate from the past to the future since these heuristics more often than not lead to successful, or at least adequate, results. The challenge for safety professionals is to identify the situations and conditions where a reliance on biased theories, models or tools might have fatal consequences. The safety consequences listed in Figure 3 can be used in safety evaluations as a checklist of potential risk factors, bearing in mind that the figure is only illustrative and not an exhaustive list of all potential effects.

One may argue that an awareness and knowledge of the phenomena discussed in this paper is first and foremost a specialist competence associated with human factors and related disciplines. Thus, the issues tackled in this paper can be considered necessary background knowledge for all human factor professionals. However, many of the phenomena discussed above are generic in the sense that more or less all (safety) management activities, at all levels, are affected by them. Therefore, we would suggest that management training, in general, should include a fair amount of information about these phenomena and their possible consequence for safety and general efficiency in organizations.

Reflection is an important aspect of the safety management process. Safety managers, developers and safety scientists should reflect from time to time on what assumptions they have concerning individuals, organizations and safety. They should also reflect on the assumptions embedded in the methods they use; do these allow systemic issues to emerge or are they biased toward some type of phenomena? This should, in the long run, also lead to enhanced quality safety management theories and methods developed in research and applied by professionals in the field. We hope that the current article provides support for a critical reflection of people's own approach to safety management.

Acknowledgements
The writing of the article was funded by the Nordic nuclear safety research (NKS) and VTT. The authors would like to acknowledge the feedback from our colleagues Elina Pietikäinen, Pia Oedewald and Ulf Kahlbom.

References

[1] ACSNI. Organising for safety. Third report of the Human Factors Study Group of the Advisory Committee on Safety in the Nuclear Industry. London: Health & Safety Commission, HMSO; 1993.
[2] Alvesson M. Understanding organizational culture. London: Sage; 2002.
[3] Antonsen S. Safety culture and the issue of power. Safety Science 2009;47:183-191.

[4] Antonsen S, Almklov P, Fenstad, J. Reducing the gap between procedures and practice – lessons from a successful safety intervention. Safety Science Monitor 2008;12:1-16.
[5] Aven T. Misconceptions of risk. Chichester: Wiley; 2010.
[6] Aven T. On how to define, understand and describe risk. Reliability Engineering and System Safety 2010;95:623-631.
[7] Aven T, Renn O. On risk defined as an event where the outcome is uncertain. Journal of Risk Research 2009;12:1-11.
[8] Barley SR, Kunda, G. Bringing work back in. Organization Science 2001;12:76-95.
[9] Bottani E, Monica L, Vignali G. Safety management systems: Performance differences between adopters and non-adopters. Safety Science 2009;47:155-162.
[10] Bourrier M. Organizing maintenance work at two American nuclear power plants. Journal of Contingencies and Crisis Management 1996;4:104-112.
[11] Cox S, Flin R. Safety culture: Philosopher's stone or man of straw? Work & Stress 1998;12:189–201.
[12] Dekker SWA. Ten questions about human error. A new view of human factors and system safety. New Jersey: Lawrence Erlbaum; 2005.
[13] Dekker S. Drift into failure. From hunting broken components to understanding complex systems. Farnham: Ashgate; 2011.
[14] Dekker S, Hollnagel E. Human factors and folk models. Cognition, Technology & Work 2004;6:79-86.
[15] Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance 1975;1:288–299.
[16] Fiske ST, Taylor SE. Social cognition. From brains to culture. McGraw-Hill; 2008.
[17] Grote G. Uncertainty management at the core of system design. Annual Reviews in Control 2004;28:267-274.
[18] Grote G. Autonomie und Kontrolle – Zur Gestaltung automatisierter und risikoreicher Systeme (Autonomy and Control – On the design of automated and high-risk systems.). Zürich: vdf Hochschulverlag; 1997.
[19] Guldenmund FW. The nature of safety culture: A review of theory and research. Safety Science 2000;34:215–257.
[20] Hackman JR, Oldham GR. Work redesign. Reading, Mass.: Addison-Wesley; 1980.
[21] Hale AR, Heming BHJ, Carthey J, Kirwan B. Modeling of safety management systems. Safety Science 1997;26:121-140.
[22] Hatch MJ, Cunliffe AL. Organization theory: Modern, symbolic and postmodern perspectives. Second Edition. Oxford: Oxford University Press; 2006.
[23] Hollnagel E. Safety management - looking back or looking forward. In E. Hollnagel, C.P. Nemeth and S. Dekker (Eds.), Resilience Engineering Perspectives, Volume 1. Remaining sensitive to the possibility of failure. Aldershot: Ashgate; 2008.
[24] Hollnagel E. The ETTO principle: Efficiency-thoroughness trade-off. Farnhamn: Ashgate; 2009.
[25] Hollnagel E, Speziali J. Study on developments in accident investigation methods: A survey of the 'State-of-the-Art'", SKI Report 2008:50, SKI 2008.
[26] Hopkins A. Lessons from Longford. The Esso gas plant explosion. Sydney: CCH; 2000.
[27] Hopkins A. What are we to make of safe behaviour programs? Safety Science 2006;44:583-597.
[28] Hutchins E. Cognition in the wild. Massachusetts: MIT Press; 1995.
[29] Johnson CW. Failure in safety-critical systems: A handbook of accident and incident reporting. Glasgow: University of Glasgow Press; 2003.
[30] Kahneman D, Slovic P, Tversky A. Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press; 1982.
[31] Katsakiori P, Sakellaropoulos G, Manatakis E. Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. Safety Science 2009;47:1007-1015.
[32] Klein G. The recognition-primed decision (RPD) model: Looking back, looking forward. In Zsambok CE, Klein G. (Eds.), Naturalistic decision making. Mahwah, NJ: Lawrence Erlbaum; 1997.
[33] Klein GA, Orasanu J, Calderwood R, Zsambok CE. (Eds.), Decision making in action. Models and methods. Norwood, NJ: Ablex Publishing; 1993.
[34] Korolija N, Lundberg J. Speaking of human factors: Emergent meanings in interviews with professional accident investigators. Safety Science 2010;48:157-165.
[35] Lundberg J, Rollenhagen C, Hollnagel E. What-You-Look-For-Is-What-You-Find – The consequences of underlying accident models in eight accident investigation manuals. Safety Science 2009;47:1297-1311.
[36] Manuele FA. On the practice of safety. Third edition. New Jersey: John Wiley & Sons; 2003.
[37] Norros L. Acting under uncertainty. The core-task analysis in ecological study of work. VTT Publications 546. Espoo: VTT; 2004.
[38] Oedewald P, Reiman T. Special characteristics of safety critical organizations. Work psychological perspective. VTT Publications 633. Espoo: VTT; 2007.
[39] Perin C. Shouldering risks. The culture of control in the nuclear power industry. New Jersey: Princeton University Press; 2005.
[40] Perrow C. Normal accidents: Living with high-risk technologies. New York: Basic Books; 1984.
[41] Rasmussen J. Risk management in a dynamic society: A modeling problem. Safety Science 1997;27:183-213.
[42] Rasmussen J. Human factors in a dynamic information society: where are we heading? Ergonomics 2000;43:869-879.
[43] Rasmussen J, Svedung I. Proactive risk management in a dynamic society. Karlstad: Swedish Rescue Services Agency; 2000.
[44] Reason J. Managing the risks of organizational accidents. Aldershot: Ashgate; 1997.
[45] Reason J. The human contribution. Unsafe acts, accidents and heroic recoveries. Farnham: Ashgate; 2008.
[46] Reiman T, Oedewald P. Assessment of complex sociotechnical systems – Theoretical issues concerning the use of organizational culture and organizational core task concepts. Safety Science 2007;45:745-768.
[47] Reiman T, Oedewald P. Evaluating safety critical organizations. Focus on the nuclear industry. Swedish Radiation Safety Authority, Research Report 2009:12. Stockholm: SSM; 2009.
[48] Reiman T, Kahlbom U, Pietikäinen E, Rollenhagen C. Nuclear Safety Culture in Finland and Sweden – Developments and Challenges. NKS-239. Roskilde: Nordic nuclear safety research NKS; 2011.
[49] Reiman T, Pietikäinen E, Oedewald P. Multilayered approach to patient safety culture. Quality and Safety in Health Care 2010;19:1-5.
[50] Robson LS, Clarke JA, Cullen K, Bielecky A, Severin C, Bigelow PL, Irvin E, Culyer A, Mahood Q. The effectiveness of occupational health and safety management system interventions: a systematic review. Safety Science 2007;45:329-353.
[51] Rollenhagen C. Att utreda olycksfall. Teori och praktik. ["Investigating accidents. Theory and practice"] Lund: Studentlitteratur; 2003.
[52] Rollenhagen C. Can focus on safety culture become an excuse for not rethinking design of technology? Safety Science 2010;48:268-278.
[53] Rollenhagen C, Westerlund J, Lundberg J, Hollnagel E. The context and habits of accident investigation practices: A study of 108 Swedish investigators. Safety Science 2010;48:859-867.
[54] Sagan SD. The limits of safety. Organizations, accidents, and nuclear weapons. New Jersey: Princeton University Press; 1993.

[55] Sawyer RK. Social emergence. Societies as complex systems. Cambridge: Cambridge University Press; 2005.
[56] Schein EH. Organizational culture and leadership. Third edition. San Francisco: Jossey-Bass; 2004.
[57] Scherer AG. Modes of explanation in organization theory. In Tsoukas H, Knudsen C, (Eds), The Oxford handbook of organization theory. Meta-theoretical perspectives. Oxford: Oxford University Press; 2003
[58] Schultz M. On studying organizational cultures. Diagnosis and understanding. Berlin: Walter de Gruyter; 1995.
[59] Smith TA. What's wrong with behavior based safety. Professional Safety, 1999.
[60] Smith P. Hits and myths. The Safety & Health Practitioner 2006;24:49-52.
[61] Snook SA. Friendly fire. The accidental shootdown of U.S. Black Hawks over Northern Iraq. New Jersey: Princeton University Press; 2000.
[62] Steele K, Pariés J. Characterisation of the variation in safety beliefs across the aviation industry. 3rd Symposium on Resilience Engineering. Juan-Les-Pins, France, October 28-30, 2008.
[63] Taleb NN. Fooled by randomness. The hidden role of chance in life and in the markets. Second Edition. New York: Random House; 2004.
[64] Teperi A-M, Leppänen A. Managers' conceptions regarding human factors in air traffic management and in airport operations. Safety Science 2011;49:438-449.
[65] Turc E, Baumard P. Can organizations really unlearn? In McInerney CR, Day RE, (Eds), Rethinking knowledge management: From knowledge objects to knowledge processes, pp. 125-146. Dordrecht, The Netherlands: Kluwer, 2007.
[66] Vaughan D. The Challenger launch decision. Chicago: University of Chicago Press; 1996.
[67] Vaughan D. The dark side of organizations: Mistake, misconduct, and disaster. Annual Review of Sociology 1999;25:271-305.
[68] Waring A. Safety management systems. London: Chapman & Hall; 1996.
[69] Waring AE, Glendon AI. Managing risk. Thomson; 1998.
[70] Weick KE. The social psychology of organizing. 2nd ed. Reading, MA: Addison-Wesley; 1979.
[71] Weick KE. Sensemaking in organizations. Thousand Oaks: Sage; 1995.
[72] Weick KE. Foresights of failure: an appreciation of Barry Turner. Journal of Contingencies and Crisis Management 1998;6:72-75.
[73] Weick KE, Sutcliffe KM. Managing the unexpected. Resilient performance in an age of uncertainty. Second Edition. San Francisco: Jossey-Bass; 2007.
[74] Wilkin P. The ideology of ergonomics. Theoretical Issues in Ergonomics Science 2010;11:230-244.
[75] Woods DD, Johannesen LJ, Cook RI, Sarter NB. Behind human error: Cognitive systems, computers, and hindsight. State-of-the-Art Report. SOAR CSERIAC 94-01. Ohio, Columbus: The Ohio State University; 1994.
[76] Woods DD, Shattuck LG. Distant supervision-local action given the potential for surprise. Cognition, Technology & Work 2000;2:242-245.
[77] Zimmermann K, Pariès J, Amalberti R, Hummerdal DH. Is the aviation industry ready for resilience? Mapping human factors assumptions across the aviation sector. In Hollnagel E, Pariès J, Woods DD, Wreathall J, (Eds), Resilience engineering in practice. A guidebook. Farnham: Ashgate; 2011.
[78] Zio E. Reliability engineering: Old problems and new challenges. Reliability Engineering and System Safety 2009;94:125– 141.