A risk management framework for a complex adaptive transport system

Jari M. Nisula

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université Toulouse 3 – Paul Sabatier*
Discipline ou spécialité : *Informatique*

Présentée et soutenue par *Jari Nisula*
Le 1 Mars 2018

Titre : *A Risk Management Framework for a Complex Adaptive Transport System*

JURY
Jean-Charles Fabre (examinateur)
Terje Aven (rapporteur)
Leslie Walls (rapporteur)
Douglas Owen (examinateur)
Philippe Palanque (directeur de thèse)
Christopher Johnson (co-encadrant)

Ecole doctorale: *Mathématiques Informatique et Télécommunications de Toulouse (MITT)*
Unité de recherche: *IRIT – UMR 5505*
Directeur(s) de Thèse: *Philippe Palanque et* Christopher Johnson
Rapporteurs: *Terje Aven et Leslie Walls*

I would like to dedicate this work to front line operators in safety critical systems: people like pilots, sea captains, train drivers and doctors. They need to take critical decisions in situations that are characterized by uncertainty, complexity and time pressure. These people inherit the latent problems in the system and must adapt and stretch to produce resilience and success on good and bad days - and nights. I count among these people my sister, an anesthesiologist working with intensive care patients, and my late father who flew a full career as an airline pilot, carrying hundreds of thousands of people safely to their destinations.

Acknowledgements

Professors Philippe Palanque and Christopher Johnson guided me through the whole academic journey during more than four years. A big thanks to them for all the help including the very valuable and record-fast feedback on the draft versions of the dissertation.

I would like to extend my thanks also to the whole team around Prof. Palanque at IRIT, in the Paul Sabatier university of Toulouse, and recognize especially the help from Dr. Célia Martinie and Dr. Camille Fayollas. The advice and support on the administrative side by Martine Labruyère and Agnès Requis was tremendously appreciated, as well as the technical support by Jean-Francois Gendet and Thierry Bichot from IRIT.

I am very grateful to the members of the Jury for their time, expertise and willingness to be part of this academic project.

The whole journey would not have taken off without the engagement of the Finnish Transport Safety Agency, Trafi. Daring to make a two-year investment to create a new unconventional approach for running their activity was bold. My thanks go to all the 70+ people in Trafi who actively participated in the project and especially to the multi-modal team of the analysis department in the heart of the project. I am particularly grateful to Mrs. Heli Koivu and DG Aviation Capt. Pekka Henttu for their invaluable support at every level and every stage of the project. Mrs. Marita Löytty deserves special thanks for helping me get access to relevant material and Mr. Ilkka Kaakinen for carrying out the final interviews and their analysis together with me.

I would like to thank Professors David D. Woods and Erik Hollnagel for the interviews and discussions I was privileged to have with them during a Eurocontrol event in Lisbon in 2014. Similarly, I am grateful for the exchanges I could have with Dr. René Amalberti in the early days of the project.

I would like to thank my friends and great professionals Dr. Simon Gill and Dr. Claus A. Andersen for their valuable feedback and encouragement.

The dissertation was written in parallel to having a full-time job. I am very grateful to Mr. Philippe Izart for providing a great amount of flexibility and thus making this challenge easier.

Finally, this long project has required many sacrifices and a lot patience from my family and especially from my wife Delphine. I have received their full support and love all this time and I am deeply grateful for that.

Table of content

List of figures

List of tables

Introduction

This work has been written for researchers and practitioners who are seeking to implement a modern risk management process for transport risks. It provides both a description of the risk management framework itself and the necessary theoretical knowledge for developing successful risk treatment strategies in the very challenging context.

Over the last ten to fifteen years, science has made significant advances in fields relevant for risk management. However, current risk management practices in industry have not yet benefitted much from these developments.

Most people associate risk with probability and severity. However, the current way to understand risk stresses the importance of uncertainty. The definition of risk by the International Organization for Standardization (ISO) might surprise many: "risk is the effect of uncertainty on objectives". According to the current view, probability is only one way to try to address uncertainty, and often not the recommended one. The result of a risk assessment should include the assumptions and the associated uncertainties. The so-called Strength of Knowledge becomes an important concept and results of risk assessment are comparable only as far as the underlying assumptions are identical. Yet another important addition to the classic understanding of risk are the so-called black swans: high-impact-low-probability events. Due to the high impact, such threats need to be addressed even when traditional risk management would ignore them due to the low probabilities. All in all, understanding of risk and risk management has evolved significantly.

Another significant evolution is the one where the dominating Cartesian world-view with the notion that like for machines, the functioning of any system can be understood as a sum of its parts, has started to give room to another very different world-view. Systems thinking and research on complexity have shown that the behavior of complex systems is dominated by a large number of interactions and influences rather than direct cause-effect relationships. Such systems feature emergent phenomena which cannot be derived from the parts, as well as surprising, counter-intuitive and non-linear behaviors. Managing risks in a complex system is a very different task from doing it in a system which is perceived fully ordered.

The need for a new type of risk management framework is clearly identified in literature. For example, Abrahamsen et al. (2004) point out that "… to support decision makers facing choices involving uncertainties about outcomes… there is no authoritative guide on how to deal with such decision problems… how the risk analysis results should be evaluated, how cost/benefit analyses should be interpreted and used, etc.". Aven (2013a) states: "it is a huge research topic to establish suitable ways of representing and treating the knowledge and surprise dimensions in risk assessment". Aven & Krohn (2014) state the following: "We need a broader concept of risk to make risk management meaningful in a black swan world, and we need to incorporate the best ideas from different traditions, including the quality management and organizational learning". Aven & Ylönen (2016) argue: "The risk assessment and management fields have developed considerably in recent years, but current industry practice, when it comes to for example the way to conduct and use risk assessments, has not changed much".

The focus of this dissertation is on transport risks, taking the perspective of a national transport safety agency, tasked with overseeing safety across several modes of transport, including aviation, maritime, railway and road safety. The research question addressed is: **What kind of risk management framework should be used for managing transport risks when the modern risk perspectives and the latest understanding of safety are embraced, and the transport system is considered a complex adaptive system?**

The dissertation is divided into three parts. In Part I, the state-of-the-art literature for the relevant topics is reviewed, covering the concept of risk and risk assessment, complex adaptive systems and safety management. Specific chapters are dedicated for risk acceptance criteria and a review of current risk management frameworks. The review of safety management includes a discussion of human limitations and biases related to risk perception and decision making. Every topic ends with a synthesis and conclusions section where the key points are summarized and the adopted paradigms are highlighted. Importantly, through the literature review, requirements for a new risk management framework (NRMF) are gradually developed.

In Part II, a NRMF in line with the developed requirements is presented. It enables risks in all transport modes to be presented in a single risk picture and supports decision-making with the aim to maximize the safety impact achievable with limited resources. The impact is further enhanced by intervention strategies such as adaptive policies and experimentation, which are well-suited to complex systems. The presentation of the framework is organized according to the ISO 31000 international standard for risk management.

Part III covers the validation of the new framework. This is done by first showing that the developed framework complies with the requirements established in Part I (full coverage). Comparison with other existing risk management frameworks shows that the developed framework goes beyond the state-of-the-art. Finally, the applicability is demonstrated with a case study, describing how components of the framework have been implemented at the Finnish transport safety agency, Trafi.

PART I – Review of the literature

Introduction to Part I

Part I covers the literature review for the topics relevant for this dissertation. The primary objective is to produce the scientific framework for designing a new improved risk management framework for a complex transport system. For this purpose, a set of requirements is gradually developed in the first six chapters. The requirements are summarized in Chapter 6.2.

The secondary objective of Part I is to provide the necessary theoretical background for people involved in the risk management process. For example, understanding the implications of the complexity (of the transport system) is essential for being able to design suitable interventions for reducing the risks. Due to this secondary objective, some topics are discussed in more length than would be necessary just for distilling the requirements for the risk management framework. This is the case especially for complexity and Safety-II.

The review is organized under several thematical chapters. Each chapter introduces the concepts, discusses them and ends with a synthesis which indicates the adopted paradigms. To make a clear distinction between what is said in the literature and the author's own comments, the latter are expressed as much as possible in the synthesis sections.

Chapter 1 introduces the context of the transport system and the risks that exist within the system. The objective of this chapter is to give the overall context of the area of application and to show concretely what the risks under study are, and what kind of accidents are caused when these risks materialize. The chapter also contains the definitions of the key terms, as well as the limitations and delimitations.

Chapter 2 introduces the concept of risk and related key concepts, such as uncertainty, strength of knowledge and the so-called black swans. This chapter lays the foundations for risk analysis, risk evaluation and risk management.

Chapter 3 covers complexity and complex adaptive systems. This chapter explains what complex systems are, what their characteristics are and what are the implications for risk management and especially for risk treatment.

Chapter 4 looks into safety management. As safety and risk are two sides of the same coin, many relevant concepts have emerged primarily in the context of safety and safety management, for instance the concept of resilience. Topics discussed in this chapter have implications for risk identification, risk analysis and risk treatment. The chapter also covers parts of human factors research relevant to risk management. Importantly, human biases and heuristics which may play important roles in risk perception and decision-making are introduced.

Only after having developed the concepts in these four chapters will it be possible to discuss the difficult topic of acceptability of risks. This is done in Chapter 5, which covers risk acceptance criteria and their ethical bases.

Chapter 6 completes the literature review. First, existing guidance for risk management frameworks is reviewed. It is then possible to establish the full list of requirements for the risk management framework. Finally, existing risk management frameworks both from industry and scientific literature are introduced.

Part I is closed with global conclusions of the literature review.

Chapter 1 – Setting the context

This chapter introduces the context of risks within the transport system. Some of the key properties of the transport system and its different modes of transport are outlined as well as the volumes involved. On the other hand, the focus will be on the risks - in other words, what can go wrong within the system and what are the consequences. To get concretely in touch with transport risks, some existing accident statistics are presented for the four modes of transport. Besides looking at worldwide statistics and statistics at the European level, a more tangible picture is drawn by taking a single country as an example and looking at transport safety statistics within that one country. This also makes sense because the scope of a transport safety agency is typically a single country. In addition to presenting the basics of transport systems, the purpose is to show why risk management within the transport system is both useful and necessary. The chapter ends with definitions and scoping of the presented research.

1 Importance of managing transport risks

The easiest way to describe the transport system is to say that it consists of everything involved with transporting people and goods. Typically, one can list four modes of transport: aviation, marine, railway and road transport. These include also light transport such as bikes and pedestrians. The system includes the infrastructure, the transport vehicles, the people, the organizations, the processes, the policies and so on. Some of the transport is commercial while a big part of it involves private people. There are also many hobbies which involve becoming part of the transport system even if the purpose is more leisure than going from one place to another: examples include sailing, parachuting and gliding. Even private drones become part of the transport system because they interfere with the rest of the system and require careful attention from the authorities and operators. The transport system is an open system so it is exposed to influences and constraints coming from the world around it. Examples of outside influences include impact of fuel prices, need for tightened security, availability of land for building new airports, pressures and regulations concerning emissions & noise and different levels of border controls and immigration procedures.

There is a commercial transport fleet of over 26,000 aircraft carrying yearly over 5 billion passengers and over 50 million tons of freight (ICAO 2015). In the EU alone, there are 880 million passengers for air travel yearly, and regional and suburban trains carry 10 times that amount being at equal level with all European metros combined, and accounting for 90% of railway passengers (ERRAC 2016). There are about 58 billion public transport journeys made in the EU yearly (UITP 2016). The worldwide commercial maritime fleet consists of roughly 90,000 vessels, carrying seaborne shipments of about 10 billion tons per year (UNCTAD 2015). It is estimated that over 90% of trade is transported by sea (Allianz 2015).

Transport acts as a key enabler and catalyst for many desirable developments. In 2015, the United Nations General Assembly adopted the resolution 70/1 introducing the 2030 agenda for sustainable development. The resolution contains 17 Sustainable Development Goals (SDGs) which split down further to 169 targets (United Nations 2015). Of these 169 targets, five are directly related to transport. Furthermore, transport is a critical enabler for at least 6 other important targets, such as access to safe drinking water, sustainable cities, reduction of food loss, agricultural productivity, air pollution and climate change mitigation (United Nations 2016).

While transport is an important part of modern societies, it also introduces risks of its own. An inherent feature of transport is moving people and goods around with a certain speed, which involves kinetic energy. This energy, often together with the presence of inflammable fuels, creates a hazard with the potential for accidents. Transport may also involve taking people to environments where they would not naturally survive, like under water or high in the atmosphere. In such cases, the safety of people is fully

dependent on the technical system providing the necessary life supporting conditions - and this introduces vulnerabilities against some specific accident types like fires.

The typical direct consequences of accidents are loss of life, injuries, environmental damage and material damages. There are, however, other potential consequences like loss of reputation, damage to a commercial brand, loss of business and social or political crises following an accident. In the following pages, some examples of accident statistics are presented. The purpose here is not to start analyzing the accidents nor drawing conclusions from the statistics but rather observe the existence of accidents in all the transport modes and appreciate their consequences. The first presented statistics are worldwide or European wide. Thereafter, in the interest of producing a more detailed and tangible example, statistics are presented concerning a single country, Finland – this also serves as a background for the case study in Part III.

In road traffic alone, over 1.2 million people die each year worldwide. In addition to the deaths, up to 50 million people incur nonfatal injuries each year as a result of road traffic crashes. As Figure 1 indicates, this makes *road* traffic crashes the main cause of death among those aged 15 to 29 years (WHO 2015).

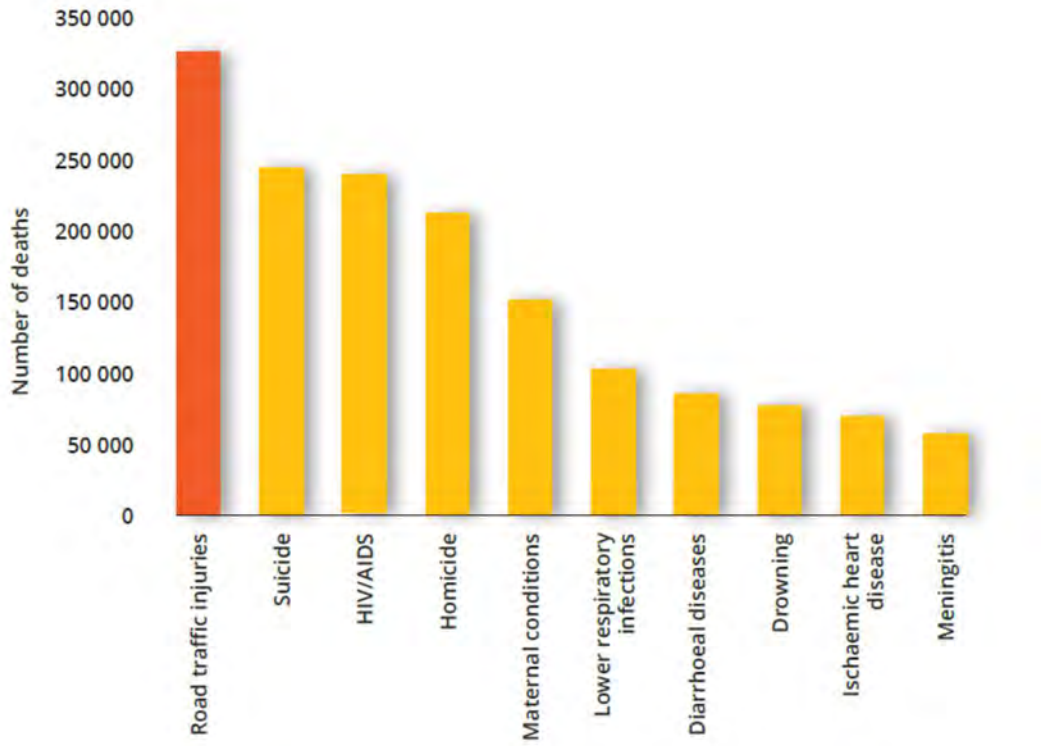


Figure 1. Top ten causes of death among people aged 15-29 years worldwide, 2012 (WHO 2015).

Figure 2 presents road traffic deaths by the type of user. It highlights the fact that there are many different types of road users each with their own safety concerns.

There are still about 1000 yearly fatalities from significant *railway* accidents within the European Union, as Figure 3 shows. The numbers in the rest of the world are typically worse, as Figure 4 shows through examples from some highly developed non-EU countries.

In the *maritime* domain, there were 2773 safety incidents ("casualties") in 2014 of which 75 were total losses (Allianz 2015). Figure 5 presents the geographical location of the total losses for the time period 2005-2014.

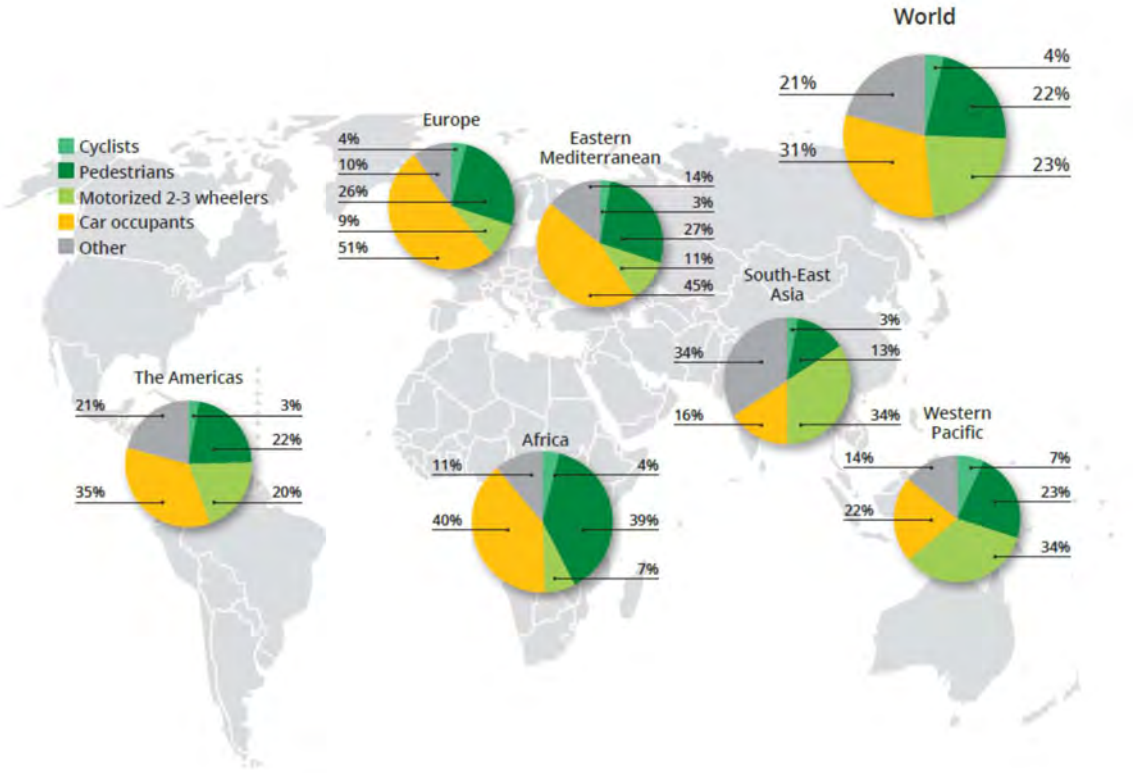


Figure 2. Road traffic deaths by type of road user (WHO 2015).

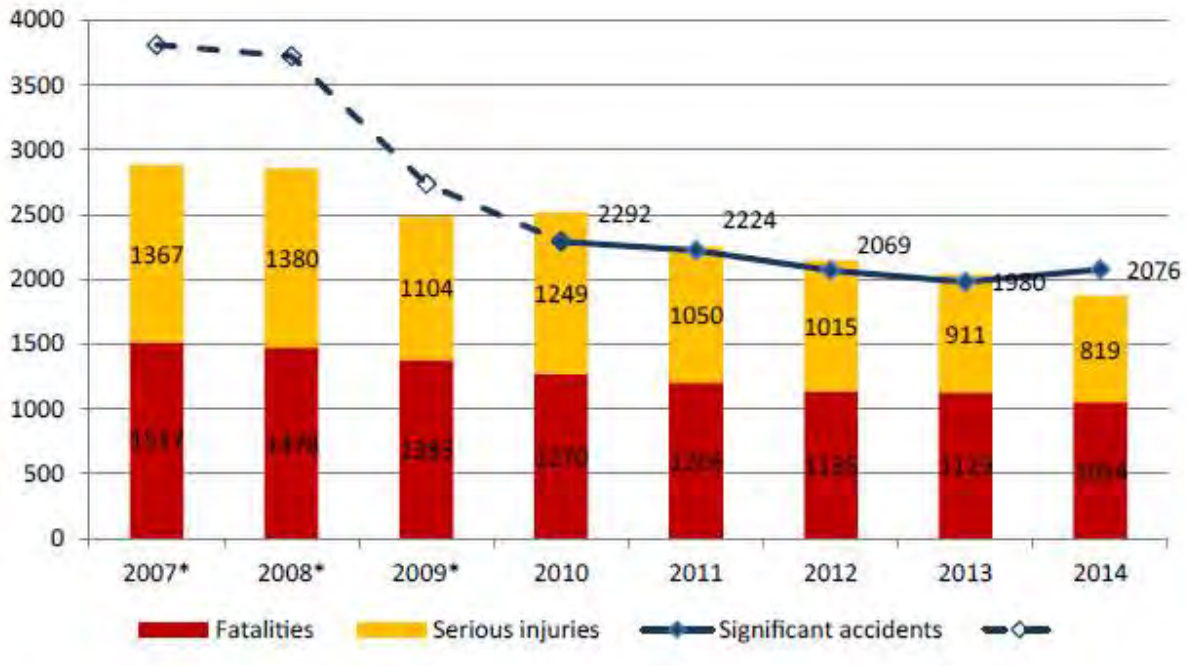


Figure 3. Significant railway accidents and resulting casualties for the EU-28 countries for 2007-2014 (ERA 2016).

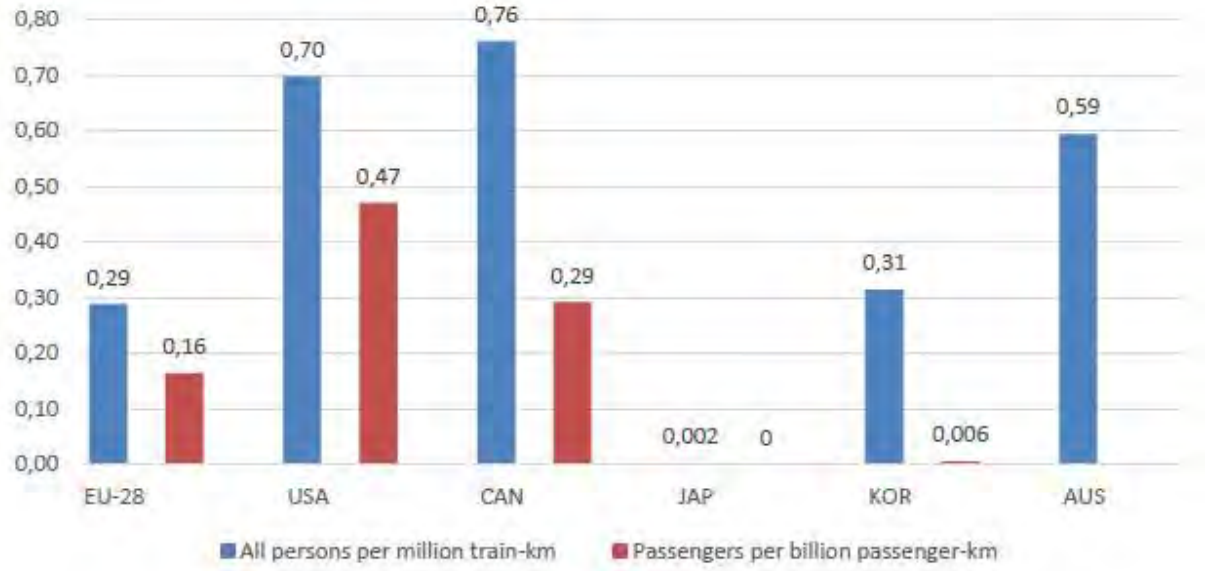


Figure 4. Railway fatality risk and passenger fatality risk for EU-28, USA, Canada, South Korea and Australia in 2010-2014 (ERA 2016).

Figure 5. Maritime total losses by top 10 regions, 2005-2014 (Allianz 2015).

Normalized by passenger kilometers, commercial *aviation* is the safest of the four transport modes (ERA 2016). Fatal commercial aviation accidents are rare but painfully visible events under strong media scrutiny. Figure 6 presents the commercial jet aviation accident rates and fatalities for 1959-2015. The fatal accident rate is less than 1 per a million departures, but the curve has become asymptotic: the rate is not improving. This is worrying as global air traffic has doubled in size once every 15 years since 1977 and is expected to continue to do so (ICAO 2016). It can also be observed that when accidents are this rare, the absolute number of fatalities vary significantly and quasi-randomly from one year to

another (influenced by the accident rate and accident types, but also by the somewhat arbitrary number of passengers on board the accident flights).

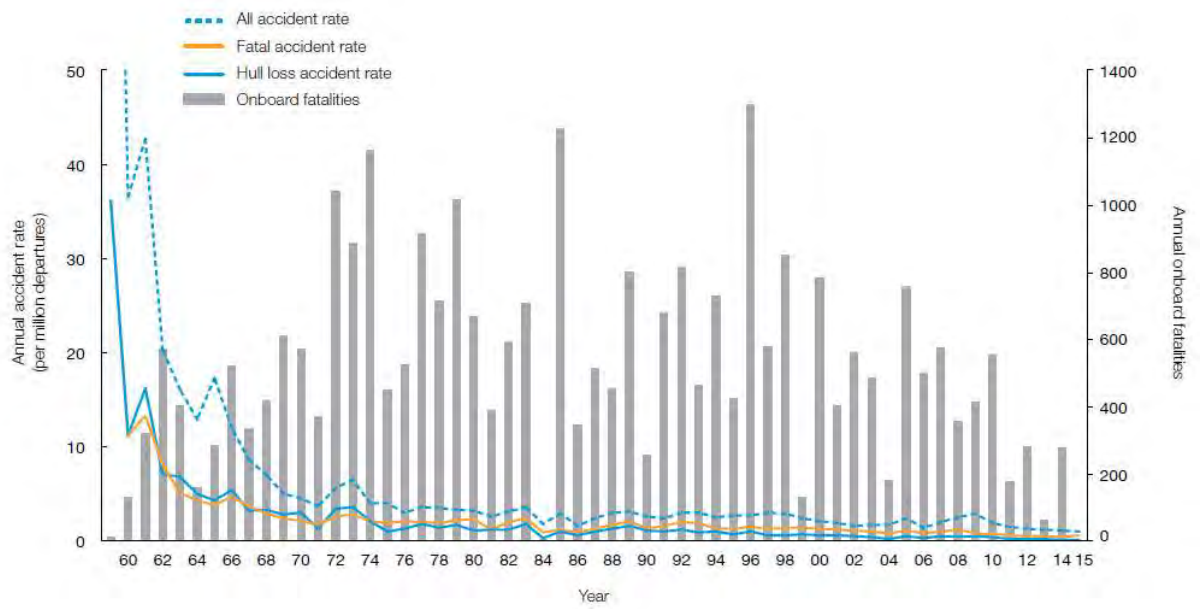


Figure 6. Worldwide commercial jet fleet accident rates and onboard fatalities by year: 1959-2015 (Boeing 2016).

The remaining figures in this chapter illustrate the reality of transport accidents in Finland. All the modes of transport are well represented in Finland due to long distances (road traffic, aviation, railways) and marine traffic both on the lakes and on the Baltic sea. Figure 7 and Figure 8 present aviation accidents in Finland in the period 2004-2014. It is worth noting that these statistics include leisure and general aviation. In the time period under study, there were no accidents within the commercial aviation category.

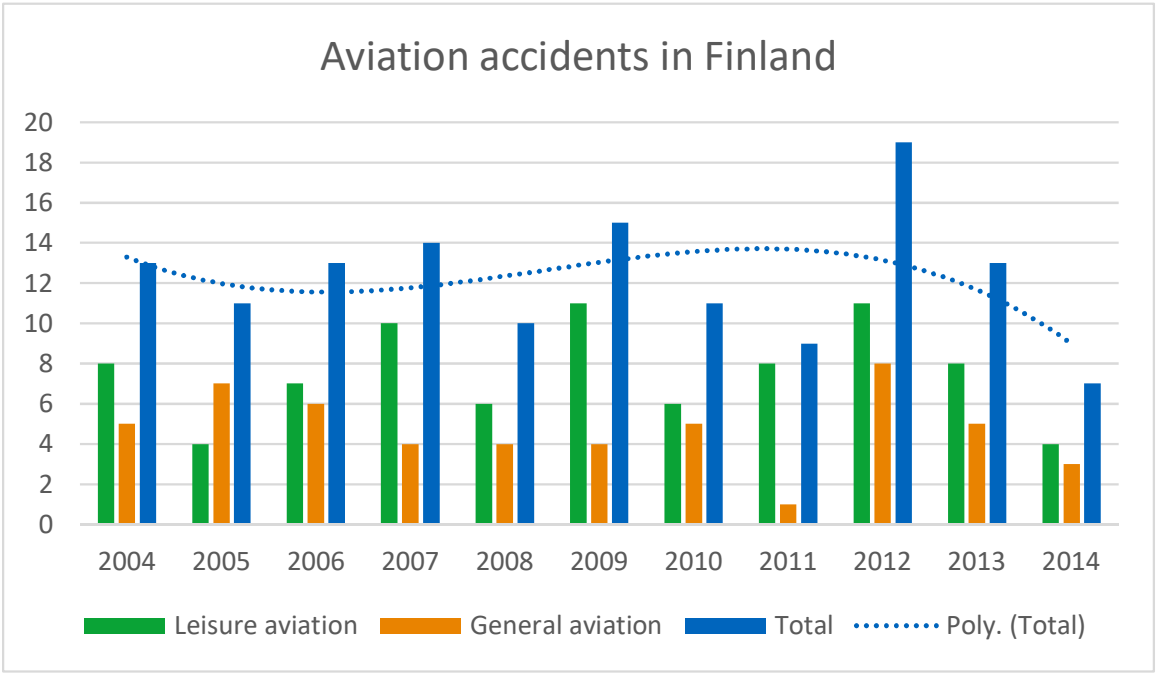


Figure 7. Aviation accidents by sector in Finland 2004-2014 (Trafi 2015a).

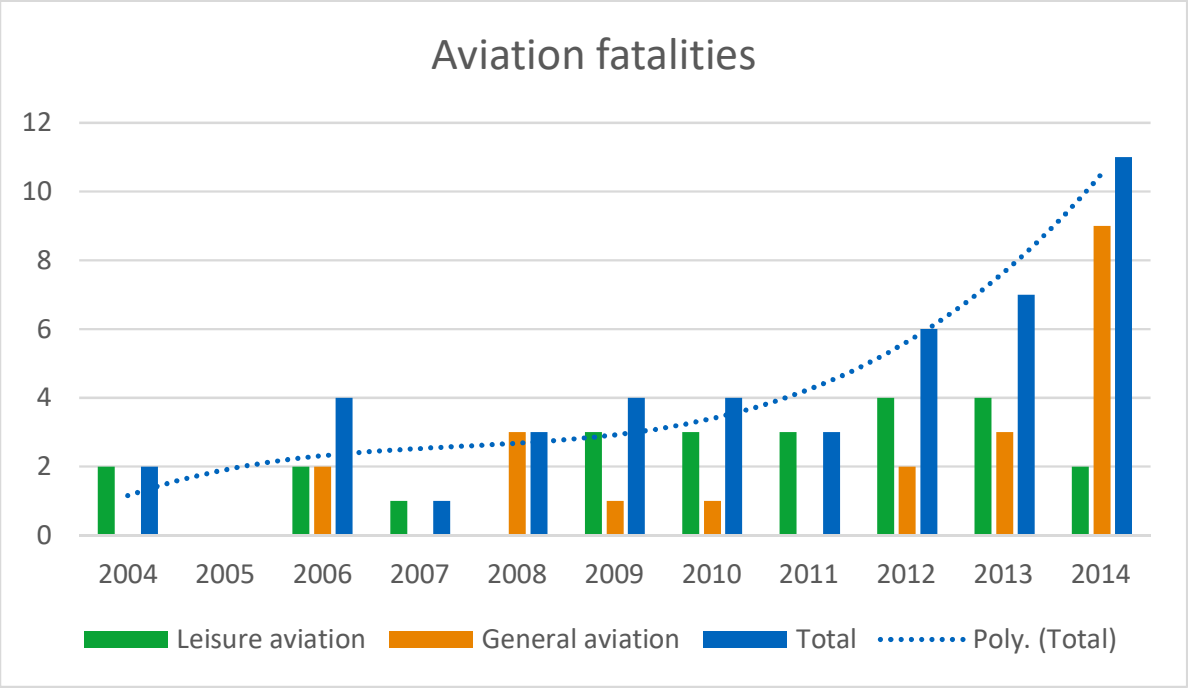


Figure 8. Fatalities in aviation by sector in Finland 2004-2014 (Trafi 2015a).

Figure 9 and Figure 10 highlight a 10-year snapshot of marine accidents/incidents in Finland both in terms of event severity and geographical location.

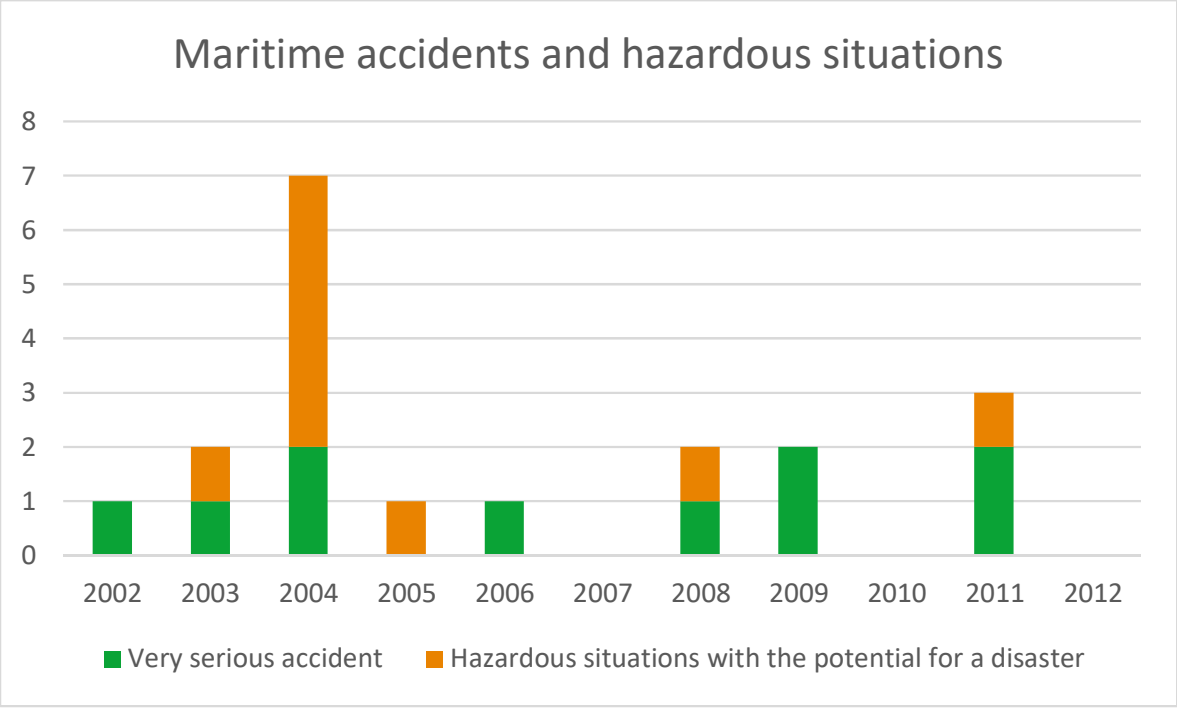


Figure 9. Maritime accidents and hazardous situations in Finland 2002-2012 (Trafi 2013a).

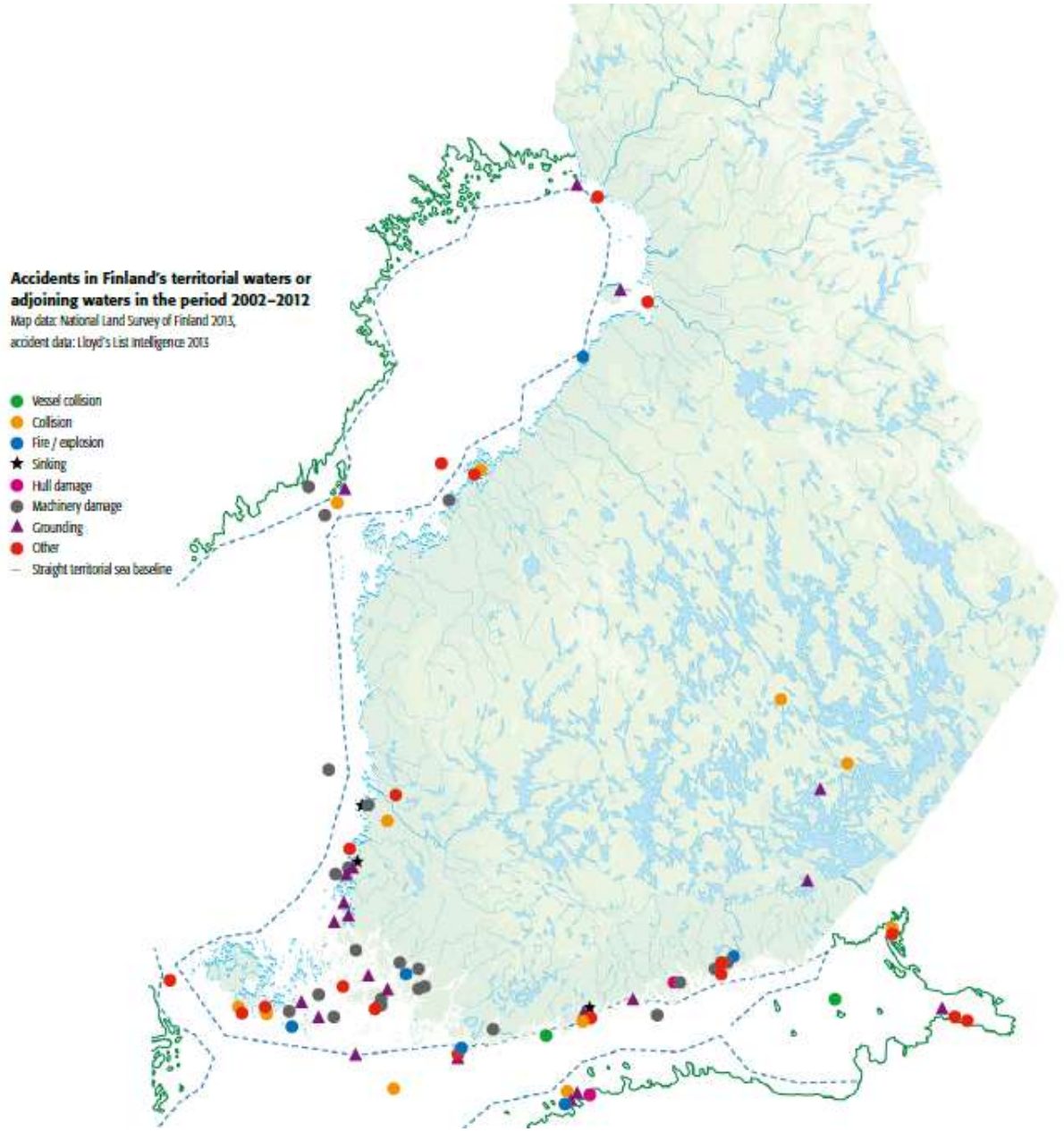


Figure 10. Maritime accidents in Finland 2002-2012 (Trafi 2013a).

Figure 11 gives an overview of rail safety in Finland. Most accidents happen at level crossings or are accidents to persons, but there are still some derailments. All in all, every year 10-25 events classified as significant accidents take place.

In 2015, there were 266 fatalities and 6385 injuries in the road traffic in Finland (Statistics Finland 2016). Figure 12 shows Finland's (FI) position among European countries and the evolution between 2010 and 2014. There is a significant difference between the best and the worst countries in Europe – a factor of five, roughly.

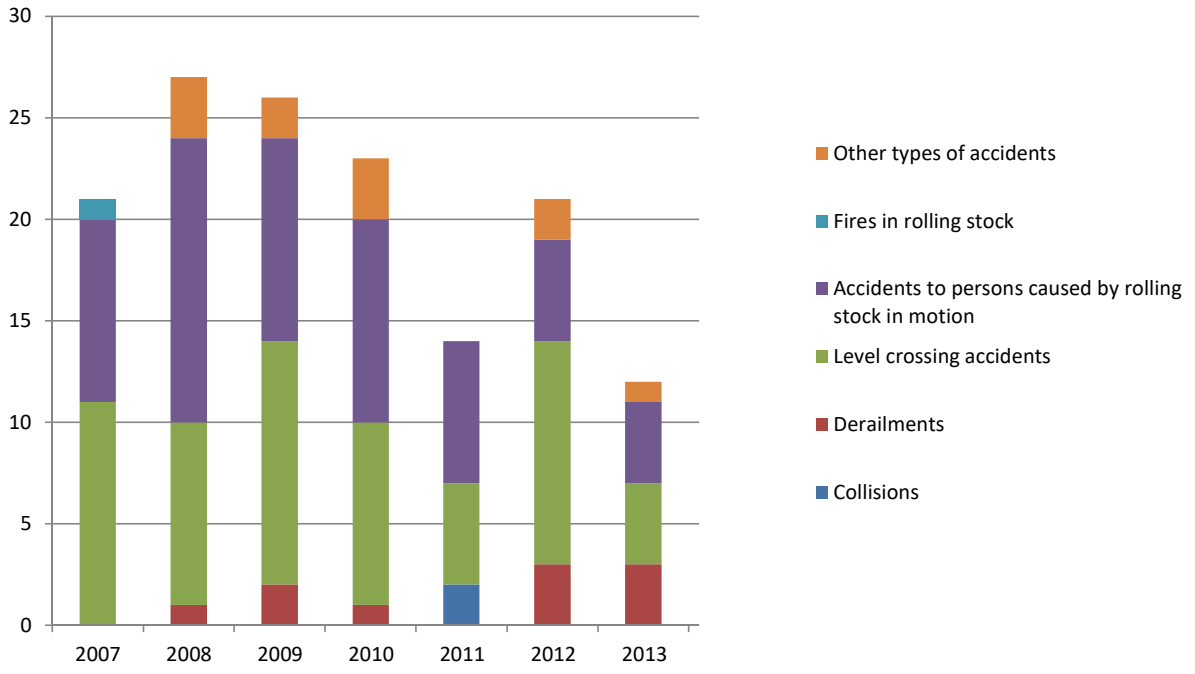


Figure 11. Significant railway accidents in Finland 2007-2013 by type of accident (Trafi 2014b)

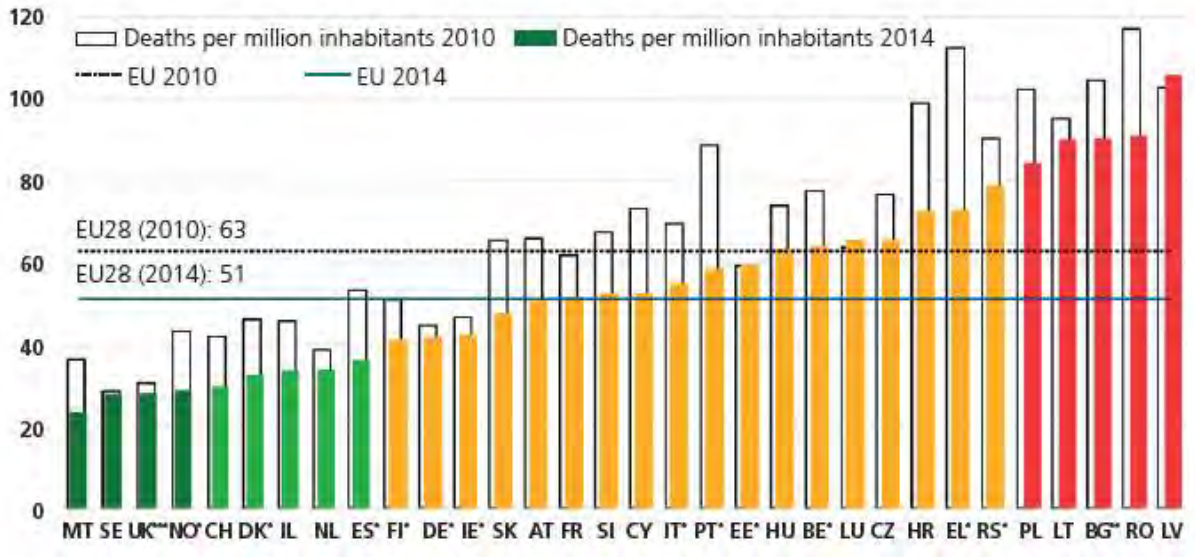


Figure 12. Road deaths per million inhabitants in 2010 and 2014 in different European countries (ETSC 2015).

The above statistics confirm that accidents occur in all four modes of transport. This means that there are real risks present in the transport system every day. On one hand, one must conclude that even though major catastrophes have become rare, their occurrence is fully possible anywhere, anytime. On the other hand, Figure 7 and Figure 8 remind that next to the commercial transport operations there is often a lively leisure activity which is visible in the transport statistics - unfortunately also in the form of accidents. This highlights the challenge of being able to allocate safety resources adequately between different types of transport-related activities, including the leisure activities. This is where a proper risk management framework must support the decision-makers.

Finally, reviewing the various accident types is also important in the sense that risk is always assessed against a reference and one needs to know what that reference is. In the case of transport risk management, the reference outcome that one wants to avoid is an accident. Having a clear understanding of this is a prerequisite for carrying out risk assessment and risk management correctly. Avoiding these accidents is the most obvious purpose of risk management within the transport system. This does not mean that accident avoidance is the *only* purpose for risk management, as it may typically have other benefits addressing more minor negative consequences and contributing to system improvement in general.

2 Definitions

The key scientific reference for risk-related definitions in this work is the Glossary published by the Society for Risk Analysis in June 2015 (Society for Risk Analysis 2015a). The 11 members of the committee behind the glossary are all recognized experts in risk analysis and their work is known through scientific peer-reviewed articles. Thanks to the committee membership, the endorsement of the Society for Risk Analysis, and the fairly recent publication date, this glossary can be considered the most credible up-to-date scientific authoritative guidance on risk analysis terminology.

Risk is naturally the most important concept in this work and the topic is discussed at length in the next chapter. Several definitions for risk exist due to its wide use in different disciplines, not to mention its flexible use in everyday language. The SRA glossary refers to *risk in relation to the consequences (effects, implication) of a future activity (interpreted in a wide sense to cover, for example, natural phenomena) with respect to something that humans value. The consequences are often seen in relation to some reference value (planned values, objectives) and the focus is often on negative, undesirable consequences. There is always at least one outcome that is considered as negative or undesirable*. The glossary then lists seven different possible qualitative definitions for risk, out of which the following can be used as the reference in this work: *risk is uncertainty about and severity of the consequences of an activity with respect to something that humans value*.

Separate from the risk concept, one can talk about the **risk description** (or metric), i.e. how does one try to describe or measure risk in practice. As discussed in the next chapter, the adopted risk description is one where *risk is described with the triplet of Consequences, measure of the associated Uncertainty and the background Knowledge that supports the two*.

The SRA glossary defines **probability** as a *measure for representing or expressing uncertainty, variation of beliefs, following the rules of probability calculus*. The more specific interpretation of probability for the purposes of the dissertation is the *subjective probability with an uncertainty standard: the probability P(A) is the number such that the uncertainty about the occurrence of A is considered equivalent (by the person assigning the probability) to the uncertainty about some standard event, e.g. drawing at random a red ball from an urn that contains P(A) x 100% red balls*.

Uncertainty can be defined both *as imperfect or incomplete information/knowledge about a hypothesis, a quantity, or the occurrence of an event*; or for a person (or a group of persons) *not knowing the true value of a quantity or the future consequences of an activity*.

Risk management is defined as *activities to handle risk, such as prevention, mitigation, adaptation or sharing*.

Risk mitigation and **risk reduction** are both defined identically as the *process of actions to reduce risk*.

The International Organization for Standardization (ISO) has also published a standard in risk management "ISO 31000 Risk management – Principles and guidelines" (ISO 2009). In comparison to the above SRA glossary definitions, the ISO standard contains the following definitions:

- **Risk** is the *effect of uncertainty on objectives*.
- **Risk Management** is defined as *coordinated activities to direct and control an organization with regard to risk*.
- **Risk management framework (RMF)** is a *set of components that provide the foundations and organizational arrangements for designing, implementing, monitoring, reviewing, and continually improving risk management throughout the organization*.
- **Risk identification** is the *process of finding, recognizing and describing risks* based on historical data, theoretical analysis, expert opinions, etc.
- **Risk analysis** is the *process to comprehend the nature of risk and to determine the level of risk*. It includes risk estimation, and provides the basis for **risk evaluation**, which is the *process of comparing the results of risk analysis with risk criteria to determine whether the risk and/or its magnitude is acceptable or tolerable*.
- The term **risk assessment** can be used to *describe the overall process of risk identification, risk analysis and risk evaluation*.
- The *process to modify risk* is called **Risk Treatment**.
- Individual *measures modifying risk* are called (risk) **Controls**. Earlier literature uses terms such as (risk/safety) barrier, defence or safeguard (Reason 1990, Maurino et al. 1995, Hollnagel 2004).
- **Consequence** is the **outcome** *of an event* affecting objectives.
- **Likelihood** is *the chance of something happening*.

This ISO standard has been criticized for not being a proper scientific framework (see e.g. Aven 2011a). In this work, the SRA glossary provides the scientific framework while the ISO terminology is used in a limited way to help organize the deliverable:

- The **New Risk Management Framework** proposed in this dissertation can be referred to with the abbreviation **NRMF**.
- The related process steps in Part II follow the ISO flow: **risk identification, analysis, evaluation** and **treatment**.
- Consequently, the term **risk treatment** is used more than risk reduction/mitigation.
- The terms **consequence** and **outcome** are used interchangeably

Some other key terms used in this dissertation include:

Components within the risk management framework are often referred to as **methods** in this dissertation. For example, the method for event risk assessment and the way to run risk workshops are both components of the risk management framework.

Loss *can be seen as a risk which has materialized*. An accident would be a typical loss. Losses are also the outcomes that one tries to avoid through risk management.

The term **accident** is generally understood as an unforeseen and unplanned event or circumstance that (1) happens unpredictably without discernible human intention or observable cause and (2) lead to loss or injury (Hollnagel 2004). The term has many different definitions within different transport domains and there is usually an attempt to give a more precise description of the level of damage involved. The

International Civil Aviation Organization ICAO (2001) gives the following definition in its Annex 13 (Aircraft Accident and Incident Investigation):

> An accident is defined as:
> An occurrence associated with the operation of an aircraft which takes place between the time any person boards the aircraft with the intention of flight until such time as all such persons have disembarked, in which:
> a) a person is fatally or seriously injured as a result of:
> - being in the aircraft, or
> - direct contact with any part of the aircraft, including parts which have become detached from the aircraft, or
> - direct exposure to jet blast
>
> except when the injuries are from natural causes, self-inflicted or inflicted by other persons, or when the injuries are to stowaways hiding outside the areas normally available to the passengers and crew; or
> b) the aircraft sustains damage or structural failure which:
> - adversely affects the structural strength, performance or flight characteristics of the aircraft, and
> - would normally require major repair or replacement of the affected component,
>
> except for engine failure or damage, when the damage is limited to the engine, its cowlings or accessories; or for damage limited to propellers, wing tips, antennas, tires, brakes, fairings, small dents or puncture holes in the aircraft skin; or
> c) the aircraft is missing or is completely inaccessible.
> Note 1.— For statistical uniformity only, an injury resulting in death within thirty days of the date of the accident is classified as a fatal injury by ICAO.
> Note 2.— An aircraft is considered to be missing when the official search has been terminated and the wreckage has not been located.

For the purpose of this dissertation, the following pragmatic definition for an accident is applied: **Accident** *is an unintended event associated with a transport operation, resulting in a loss which could include human, environmental, financial and other dimensions*.

An **incident** is defined as *an operational event which either resulted in some losses but not to the extent of being considered an accident; or had the potential to become an accident*. Again, more specific definitions would be available for different modes of transport.

Definitions for the term **safety** are discussed in Chapter 4. There is a distinction between safety and **security** - the latter addressing intentional acts of harm. In the world of safety and safety risk management, there may be known and hidden risks in the system, but nobody is intentionally trying to produce an accident.

The term **safety event** can be used *for any operational event which could be considered to have safety implications*, irrespective of whether it reaches the defined thresholds for becoming an incident or an accident in the official sense.

In the context of mandatory reporting of safety events to the authorities, the terms **occurrence** and mandatory occurrence reporting are often used. In this respect, occurrences are simply *safety events which fulfill the criteria for being officially reported to the authorities*.

In the context of safety events, losses and risk, it is important to highlight the difference between actual and potential outcomes. The **actual outcomes** of an event are the *tangible outcomes of the event in the real world as the event unfolded*. The **potential outcomes** refer to the *outcomes that could have resulted in if the event had escalated and become an accident* (or an incident). For example, in a near-collision between two large passenger aircraft the actual outcome is ”nothing”, but the potential outcome is a

midair collision with several hundred fatalities. The term potential outcome is often used in the context of a specific scenario.

Severity refers to the *magnitude of loss (damage, harm, etc.)*. In the context of risk assessment, severity usually refers to a scenario (of loss) which is possible but which has not yet materialized. Therefore, in this context the focus is on a **potential severity** associated with the scenario.

In this dissertation, there is a hierarchy between three important terms related to safety and accidents: hazard, safety issue and scenario. A hazard is the most generic of these three terms. The definition of **hazard** can be based on the following SRA glossary definitions:

- **Harm** is physical or psychological injury or damage
- **Damage** is loss of something desirable.
- **Risk source** is an element (action, sub-activity, component, system, event, etc.) which alone or in combination with other elements has the potential to give rise to some specified (typically undesirable) consequences.
- **Hazard** is a risk source where the potential consequences relate to harm.

For example, a strong gusty wind is a hazard. Inflammable fuel is another hazard. A **safety issue** is defined as a *specific concern about a certain type of accident risk in a specific context*. It can be seen as a hazard or a combination of hazards in a specific defined context. For example, an operator may consider its operation to a specific airport a safety issue due to specific hazards associated with that operation (frequent strong gusty crosswind, frequent bird activity, short runway, etc.). Typically, a safety issue may be identified due to one or more safety events. For example, fatigue reports emerging repeatedly from a specific route may lead to the identification of the related safety issue. A **scenario** is a more *detailed description of how a specific accident type can develop within a given safety issue*. A single safety issue may contain several scenarios. For example, incorrect loading of an aircraft may be identified as a safety issue. That safety issue could contain several scenarios: a) loss of control due to *unfastened* cargo moving during the flight; b) loss of control due to *overweight*; c) loss of control due to *center of gravity* out of limits. Usually a scenario is the level at which risk can be assessed, i.e. one cannot perform risk assessment on a hazard.

Safety issues and scenarios can also be called **threats**. In other words, the meaning given to threat in this dissertation is slightly more specific than that of hazard. The term threat can be useful as for different people the term "risk" could mean both the "situation or event where something of human value is at stake and where the outcome is uncertain" and the "expected value/probability" (Aven 2014a, pp. 230-231) i.e. the magnitude of the risk. One can say "there is a risk of a structural failure here"; and one can ask, "how high is the risk related to investment A?" The term threat then refers only to the first meaning, i.e. *the safety issue/scenario meaning of the word*.

A **risk picture** *is a visual presentation where several risks are presented simultaneously so that their characteristics and importance can be compared*. Ideally, a risk picture helps to gain an overall understanding of safety and risks within a certain system or operation and thereby supports prioritization and decision making. Based on the above definition of the term threat, the risk picture could also be called "picture of threats".

The terms data and information are so widely used in everyday language that it is difficult to give them any specific meanings. A generic scientific definition is developed in Chapter 3.5.3 by introducing the Data-Information-Knowledge-Wisdom (DIKW) framework. In the context of transport safety, the term (safety) **data** can also be *used for the safety information which is collected from the operations and arrives to the safety agency as an input to the risk management process*. In the most typical case, this data comes from an operator and the operator has already processed the raw data to a more refined format through the use of keywords, categories, etc. An occurrence report would be the most typical example. **Raw data** *is safety data in its initial state as received from the initiator* (human or technical) before the data is processed in any way. In the same context, **Safety information** *can be understood as*

a more generic term and can be used - for example - for referring to an awareness of current operational conditions and practices, as well as overall information about a particular operator and its operational and safety practices.

3 Limitations and delimitations

The chosen topic implies two main limitations, one related to the concept of risk and the other stemming from the fact that the transportation system is already extremely safe.

Risk is an abstract concept. Unlike accidents which have measurable and quantifiable real-world consequences, risk cannot be touched nor measured in the real physical world. Risk only exists in the minds of people. Therefore, by definition, it is subjective. There is no doubt that the chosen methodology for risk assessment will have an impact on the level of subjectivity in the result and one can try to reduce subjectivity through appropriate design of the methodology. The important remaining limitation is that the final correct answer is not available anywhere. Any result related to risk assessment can always be challenged and no authority possesses the right answer.

Ironically, the context of dealing with very safe systems leads to another limitation which has implications on the validation of the work. It is common practice to relate the safety of a transport system to how many accidents and fatalities it suffers compared to the volume of traffic (see e.g. International Transport Forum 2015, p. 14, where "road safety performance" is illustrated by statistics on road fatalities per inhabitants and vehicle-kilometers; or Trafi 2013b, p. 16, where the "status of aviation safety" is presented through statistics on accidents per 10.000 flight hours). Unsafe systems would have a lot of accidents, and safety improvements in such systems could be validated by observing the resulting reduction in accidents. In very safe systems however, there are already hardly any accidents. Therefore, accident data cannot be used to validate any safety improvement. Amalberti (2001) has discussed this in length and introduced the term ultra-safe systems. He notes that the time duration necessary for seeing the effect of a safety measure in accident statistics becomes excessively long (Amalberti (1999) quotes a delay of four years for a system with a 10^{-5} accident rate and about six years with a 10^{-7} accident rate). Attributing improvement to a specific measure becomes virtually impossible due to the thousands of changes the system would experience in such a time period. As will be seen, complex systems are particularly intractable in this respect. In road traffic where there is a considerable number of fatalities each year, improvements should be more quantifiable. However, even there it would be difficult to prove what the factors behind the improvements are.

The main delimitation is that the focus is on safety in the classic sense, i.e. *malicious acts* are outside the scope. This means that the risk management framework is developed for *safety* risks, not for security risks. The dynamics change significantly when somebody is actively trying to breach the barriers and to maximize losses. Despite this delimitation, security is kept in mind as an important constraint and priority which needs to be considered when interventions are designed, especially that safety and security concerns are often in conflict: e.g. securitywise, cockpit doors should be unbreakable while safetywise they should be easy to open in case of evacuation or incapacitation of a pilot.

Chapter 2 – Current perspective to the concept of risk

The objective of this chapter is to gradually develop the different aspects of the modern understanding of risk, so that the new proposed risk management framework (NRMF) is based on this solid foundation.

For the past decades, the most common way to understand risk has been to consider it as a combination of *probabilities* and *consequences*. For example, the nuclear industry which developed detailed methodologies for system safety analyses, adopted a quantitative definition of risk, where each scenario s_i is linked with the corresponding probability p_i and associated consequences c_i (Hafver et al. 2015, Jore & Egeli 2015). Such an approach to risk can be associated with the so-called *classical positivist* view on risk. The positivist view sees risks as something that exists objectively and the analysis as a tool to estimate the objective risk. In the positivist worldview, risk consists of physical facts that can be explained, predicted and controlled. The obtained values for risk are then considered objectively true and free of value judgments (Jore & Egeli 2015). The practical methods used are based on probabilities and expected values.

Anyone having worked in industry during the last few decades and having been exposed to safety work or risk management can confirm the dominance of this very simple view of risk as an expected value, dealing with probability and severity of consequences. Aven & Ylönen (2016) note that while the risk assessment and management fields have developed considerably in recent years, the current industry practice in terms of risk assessments has not changed very much from the times when the tools were developed in the 70s and 80s. This can be verified by reviewing some of the guidance documents and method reviews related to risk management within different industries. Examples can be found in ICAO (2013, p. 5-20), SMICG (2013), FAA (2014, p. 34), Khan et al. (2015, p. 123), IMO (2006, p.2), Montewka et al. (2014, p. 79), RSSB (2009, p. 12).

A more recent view on risk is the so-called *constructivist* perspective. This view is based on the finding that the probability-based perspective on risk is too narrow and it does not adequately address the *uncertainty* related to different possible consequences. According to this view, probability is only one way to describe uncertainty and the concept of risk must take into account other important aspects such as the knowledge dimension, the unforeseen and potential surprises (Aven 2013, Aven & Krohn 2014, Jore & Egeli 2015). It is argued that risks expressed as expected values do not reflect the knowledge in which they are based. The so-called strength-of-knowledge (SoK) dimension becomes a key aspect of risk (Aven & Krohn 2014). For example, consider two scenarios which have the same potential negative consequence and the assigned (subjective) probability of this consequence is the same in both scenarios. However, it may be that in one case there is high level of certainty about the probability, whereas in the second scenario the strength of knowledge is very weak. From the classical positivist point of view the two scenarios would have the same risk. However, from the constructivist perspective the two cases would be different because of the different levels of uncertainty. Aven (2012) proposes a way to denote risk along this modern view as (C, U) where C denotes the future consequences of the activity and U expresses that C is unknown. Another symbol A can be added to denote the hazards/threats: (A, C, U). One can also stress the *knowledge* dimension by describing risk as (C, Q, K) where Q is a measure of *uncertainty*, and K is the background knowledge that C and Q are based on (Aven & Krohn 2014). It is worth noting that the current official international definition of risk by the International Organization for Standardization (ISO) is also based on the concept of uncertainty: risk is defined as *effect of uncertainty on objectives* (ISO 2009).

What is common to the different views on risk is that, first of all, the focus is on the *future*, and secondly one tries to assess the potential negative impact of one or more scenarios. This negative impact is on something that humans value, typically harm to humans, the environment or material assets. One may

also consider effects such as loss of revenue and social disruption. The harm on humans could include death, injury, permanent disability, or illness. These could affect workers, users, members of the public, or even unborn generations (Johansen & Rausand 2014).

Abrahamsen et al. (2004) argue that risk analyses are always based on background information - even simple calculations of expected values. Difficult ethical and political deliberations involving complex value judgments cannot be achieved through a simple mathematical formula. The background information must always be reviewed with the proposed results. With the above considerations, it becomes obvious that representing risk with a simple number would be very undesirable as this would not at all reflect the associated assumptions and value judgments (Montewka et al. 2014). One compromise between a simple number and very comprehensive risk descriptions is to use a vector consisting of the outcome component and the probability component (Johansen & Rausand 2014).

Aven (2012) summarizes the discussion on risk definitions by saying that there is no agreed definition of the concept of risk. Literature includes a number of different ways of understanding the risk concept. He also highlights two important characteristics of a risk definition. First, risk should be defined in such a way that there is a distinction between risk per se and *how risk is managed*. Second, the risk definition should allow distinguishing between risk per se and *how risk is perceived*.

Montewka et al. (2014) propose a terminology where *risk concept* concerns "what risk means in itself" i.e. "what risk is"; a *risk perspective* is "a way to describe risk, analyze and make statements about risk"; and a *risk metric* is "the assignment of numerical value to an aspect of risk according to a certain standard or rule".

Despite the lack of a universal agreement on the definition of risk, it is clear that the modern risk perspective, focusing on uncertainty, knowledge and surprises, is gaining ground. The Specialty Group on Foundational Issues in Risk Analysis of the SRA, consisting of nine experienced and active researchers, has published the document "Risk Analysis Foundations" (SRA 2015b). The importance of both the knowledge dimension as such and surprises is highlighted, and the concept of Strength of Knowledge is promoted as a broader way of expressing the uncertainties than the probability assignments alone. Both the SRA Glossary (2015a) and the ISO definition are in line with the modern view and there are even the first signs of industry picking up the new risk perspectives (Aven & Ylönen 2016). Aven (2013a) summarizes the so-called new risk perspective as the sum of:

- Probability-based thinking
- Knowledge dimension
- Surprises (black swans)

These three aspects are discussed in the following three sections. The fourth section introduces the concept of risk aversion as it will be a recurring topic in Part I.

1 Clarifying the concept of probability

Aven (2012) points out that many different risk definitions and interpretations are closely related to different understandings of *probability*. There is still substantial discussion within the scientific community around the meaning and interpretation of probability. The *objectivist* (or frequentist) view is that probability is primarily a physical characteristic reflected in the relative frequency within a repeatable experiment, e.g. when throwing the dice. According to the *subjectivist* view, probability is a *personal belief* and it is never possible to know probabilities for certain. Aven & Reniers (2013) bring up the fact that it is often very difficult to apply the (frequentist) concept of repeatable experiments in the real world. Can, for example, different oil drilling platforms be considered "similar" in the sense that the operation on each one of them together build up the "population" and an accident probability could be interpreted as a number of times the "experiment" of safe operation fails in this population? Or, if the average accident rate on a commercial flight is one in 5 million, can all these flights be considered similar enough to interpret the accident probability for a given flight as one in 5 million?

Paradoxically, the individual oil drilling platforms or the individual flights should be similar enough to allow this interpretation, but not too similar either, because otherwise they would all either have an accident or not. As Aven notes, it is difficult to specify what should be the (fixed) framework conditions for the situations in the population and what could be varying.

The subjectivist view can be completed with an uncertainty standard, i.e. a concrete example of uncertainty given as a reference. For example, if a person assigns the probability of 0.1 for an event E, the person compares the uncertainty (or degree of belief) of E occurring, with drawing a specific ball from an urn containing 10 balls. Aven & Reniers (2013) argue that if the separation is required between uncertainty assessments per se and value judgments, there is only one universally applicable probability concept. It is the subjective probability with a reference to an uncertainty standard. They also note that the term *subjective* probability may sound non-scientific, and therefore it is sometimes replaced by other terms such as *judgmental* or *knowledge-based* probability. In this dissertation, probability is understood as the subjective probability, unless when specifically stated otherwise.

2 The knowledge dimension of risk

In real systems, uncertainty is a key aspect of risk assessment. According to Bjerga & Aven (2015), the uncertainty is related to potential future events and consequences and is due to limitations in knowledge. There is uncertainty both on the *system model* and the *input parameters*, either because these are not known or there is no shared view on these between the stakeholders. The real-life system cannot be characterized exactly because the knowledge of the underlying phenomena is incomplete (Aven & Zio 2011). According to Paté-Cornell (2002), uncertainties can be divided into two categories: epistemic and aleatory. The former result from the lack of fundamental knowledge while the latter reflect randomness in a well-defined statistical sample.

Makridakis et al. (2009) list empirical evidence related to future uncertainty. The list contains, for example:

- The future is never exactly like the past. The extrapolation of past patterns or relationships cannot provide accurate predictions.
- Statistically sophisticated, or complex, models fit past data well but do not necessarily predict the future accurately.
- Both statistical models and human judgment have been unable to capture the full extent of future uncertainty.
- Forecasts made by experts are no more accurate than those of knowledgeable individuals.

These points cast a serious doubt on the capability to reduce uncertainty in risk assessment through system modellization.

Complex systems will be the topic of Chapter 3. In complex systems, there is yet another dimension of uncertainty: the nature of the system is such that it does not allow the full understanding of the system logic nor reliable prediction of the system behavior (Bjerga et al. 2016). There is a fundamental shift here from “unknown” to “unknowable”. Like Walker et al. (2010) state: “Most of the important policy problems currently faced by policymakers are characterized by high levels of uncertainty, which cannot be reduced by gathering more information. The uncertainties are unknowable at the present time, but will be reduced over time.”

In addition to uncertainty related to lack of knowledge and the impossibility to model intractable (complex) systems, there are also fundamental surprises, “black swans” which will be discussed in Chapter 2.3. In conclusion, people engaged in risk management will have to learn to deal with uncertainty.

Acknowledging the role of uncertainty in risk assessment entails on one hand dealing with the uncertainty during the risk assessment, and on the other hand, communicating on the uncertainties in

relation to the results obtained. As Abrahamsen et al. (2004) note, *risk analyses are based on background information that must be reviewed together with the results*.

In terms of dealing with the uncertainty during risk assessment, two topics remain to be discussed. First, several ways to assess (and possibly categorize) the level of uncertainty have been proposed. The aim here is to gain an understanding on how much uncertainty the risk assessment contains and to communicate that in an understandable and comparable way together with the results of the risk assessment. Secondly, the main objective of the risk assessment still remains: there needs to be a way to obtain useful results despite the uncertainty. The following paragraphs address these two topics.

Montewka et al. (2014) refer to knowledge on model parameters as *knowledge* (K) and understanding of the system behavior as *understanding* (N). They adopt the symbology:

$\Delta \sim \{K, N\}$

Where Δ represents the set of knowledge and understanding dimensions used within the risk model. They formulate risk as:

$R \sim \{A, C, Q\}|\Delta$

Where A denotes events, C consequences, Q uncertainty analysis (in line with Aven & Zio 2011) and the $|\Delta$ shows that all elements of the risk description are conditional upon knowledge and understanding. Montewka et al. (2014) also propose a simple qualitative uncertainty assessment method. Quality of knowledge (K) is graded on a scale *good-moderate-poor* and level of understanding (N) is graded on a scale *high-medium-low*. Some descriptors are given for each category to guide the assessment process. A classification table is given for defining the final result (L-M-H) based on these two components. For example, low knowledge together with medium understanding would result in a medium uncertainty level. Montewka et al. combine the obtained parameter *uncertainty* with parameter *sensitivity* to derive what they call parameter *importance*. The idea is, that increasing uncertainty combined with high sensitivity means that the parameter has a high uncertainty impact on the result of the risk assessment. On the other hand, high uncertainty of a parameter may not be so important if the sensitivity of the risk assessment for this parameter is very low. This presented method provides a way to present the perceived uncertainties in a systematic way and point out the most critical parameters, which may deserve further investigation.

Walker et al. (2010) present a classification of uncertainty into four levels. In summary:

- Level I uncertainty can be described adequately in statistical terms.
- Level II uncertainty implies that there are alternative trend-based futures and/or different parameterizations of the system model. Level II uncertainty is often captured in the form of a few trend-based scenarios based on alternative assumptions. The scenarios are then ranked accordingly to their likelihood.
- Level III uncertainty represents deep uncertainty about the mechanisms and functional relationships being studied. There is little scientific basis for placing believable probabilities on scenarios. Level III uncertainty is often captured in the form of a wide range of plausible scenarios.
- Level IV uncertainty implies the deepest level of recognized uncertainty. In this case we only know that we do not know.

Broadly speaking, levels I and II correspond with what Makridakis et al. (2009) call “Subway uncertainty”, and levels III and IV correspond with “Coconut uncertainty” (Walker et al. 2010). The word picture of subway uncertainty relates with statistical variance in the time a given someone arrives to work using the subway. The word picture of coconut uncertainty refers to the very rare but possible occasion where a coconut drops on a tourist’s head causing a fatality. Aven (2013b) notes that at levels I and II, uncertainties can be represented by probabilities, and at levels III and IV this is not possible.

Paté-Cornell (2002) argues that the use of expert judgment is simply unavoidable. One can try to increase the strength of knowledge to the practical maximum by collecting as much of the relevant information as feasible. For the rest, all assumptions have to be described clearly and results can be compared only to the extent that they have been computed on comparable bases (e.g. if different scenarios have been risk assessed with different levels of conservatism in the assumptions, the related results are not comparable).

Walker et al. (2010) refer to *deep uncertainty* as "the condition in which analysts do not know or the parties to a decision cannot agree upon:

- the appropriate models to describe interactions among a system's variables
- the probability distributions to represent uncertainty about key parameters in the models
- and/or how to value the desirability of alternative outcomes."

Cox (2012) characterizes deep uncertainty in the following way:

- Well-validated trustworthy risk models giving the probabilities of future consequences for alternative present decisions are not available,
- The relevance of past data for predicting future outcomes is in doubt,
- Experts disagree about the probable consequences of alternative policies - or, worse, reach an unwarranted consensus that replaces acknowledgment of uncertainties and information gaps with groupthink, and
- Policymakers (and probably various political constituencies) are divided about what actions to take to reduce risks and increase benefits.

Aven (2013b) builds on the uncertainty taxonomy presented by Walker et al. (2010) and presents an alternative uncertainty classification taxonomy where the *strength of knowledge* is linked with the possible occurrence of black swan events. He distinguishes between three levels of uncertainties: low moderate and deep. He argues that:

- low uncertainties are associated with strong knowledge and no black swans;
- moderate uncertainties are associated with some dominating explanations and beliefs and the possibility that a black swan may occur;
- deep uncertainties are associated with poor knowledge and no black swans: it is argued that it is not meaningful to refer to black swans when the knowledge is poor.

Aven (2013a) argues that when dealing with uncertainty related to risk assessment, the *strength of knowledge* concept reflects more precisely the right ideas than the uncertainty concept itself. He proposes two methods for grading the strength of knowledge. The first one, based on Flage & Aven (2009), uses a weak-medium-strong scale for SoK. The knowledge is considered *weak* if one or more of these conditions are true:

a) The assumptions made represent strong simplifications
b) Data are not available or are unreliable
c) There is lack of agreement/consensus among experts
d) The phenomena involved are not well understood; models are nonexistent or known/believed to give poor predictions.

If all of the following conditions are met, the knowledge is considered *strong*:

- The assumptions made are seen as very reasonable
- Much reliable data are available
- There is broad agreement/consensus among experts
- The phenomena involved are well understood; the models used are known to give predictions with the required accuracy.

Cases in between are classified as having *medium* strength of knowledge.

The second method for assessing the strength of knowledge in Aven (2013a) starts by the identification of all the main assumptions behind the risk assessment. These assumptions are converted to a set of

uncertainty factors. This is followed by an analysis on the impact of deviations related to these uncertainty factors, in order to determine the criticality or importance of each assumption. This measure is called *assumption deviation risk*. It reflects the deviation from the assumption and the associated consequences, measure of uncertainty of this deviation, and the knowledge that these are based on. The idea is, that a high overall uncertainty value will have a direct impact on the result of the risk assessment – either by upgrading the risk level or by encouraging risk reduction measures even in cases where the pure cost-benefit/cost-effectiveness analyses would not justify them.

The preceding paragraphs illustrate both the key role of uncertainty in risk assessment and the struggle to quantify it one way or another. This becomes necessary as the strength of knowledge now becomes one of the key parameters in risk assessment.

There is an interesting parallel between integrating the knowledge dimension in risk analysis and the so-called NUSAP system. NUSAP is a notational system for the management and communication of uncertainty in science for policy, first introduced by Funtowicz & Ravetz (1990). The letters in NUSAP stand for (Sluijs et al 2005):

- Numeral: the number expressing the magnitude.
- Unit: the unit that goes with the numeral.
- Spread: expresses the variation in the result, this could be based on statistical data analysis, sensitivity analysis, Monte Carlo analysis, possibly in combination with expert elicitation.
- Assessment: qualitative judgments about the information.
- Pedigree: evaluative account of the production process of information, using qualitative expert judgment and a pedigree matrix.

The aim of NUSAP is to capture both quantitative and qualitative aspects of uncertainty and communicate them in a standardized way. In this respect, spread, assessment and pedigree are similar to the Strength of Knowledge dimension in risk assessment, adding the knowledge dimension beyond the simple numeral value. For a discussion on the exploration of similarities and synergies between NUSAP and uncertainty-based risk assessment, see Berner & Flage (2016).

3 The need to address black swans

Although the concept of black Swan has existed for a long time, Nassim Nicholas Taleb introduced the concept in its current – risk management related - meaning in his book “the black swan”, first published in 2007. He described a black swan event as something which:

- is very surprising, an outlier, outside the realm of regular expectations because nothing in the past would point to its possibility.
- has an extreme impact
- is typically rationalized and explained after the fact as something which was predictable (Aven 2015).

Aven (2013d) takes on the challenge of putting Taleb’s black swan concept into a scientific frame. He summarizes the meaning of the black swan as a *surprising extreme event relative to the present knowledge/beliefs*. Later Aven & Krohn (2014) and Aven (2015) present a more refined definition where three different types of black swan events are identified:

a) events that were completely unknown to the scientific environment (unknown unknowns)
b) events not on the list of known events from the perspective of those who carried out the risk analysis, but known to others (unknown knowns)
c) events on the list of known events in the risk analysis but judged to have negligible probability of occurrence, and thus not believed to occur.

The following list from the 2010 edition of his book, further illustrates Taleb’s views about black swans:

- Black swan logic makes *what you don’t know* more relevant than *what you know* (p. xxiii).

- The Gaussian bell curve cannot handle large deviations and therefore ignores them, yet makes us confident that we have tamed uncertainty (p. xxix).
- Humans have a pathology of thinking that the world in which we live is more understandable, more explainable, and therefore more predictable than it actually is (p. 9).
- *Epistemic arrogance* has a double effect: we overestimate what we know, and underestimate uncertainty (p. 140).
- The terms *mediocristan* and *extremistan*, coined by Taleb, where the former refers to linear, scalable phenomena – and the latter, to a very non-linear system, where a single observation can disproportionately impact the aggregate, or the total (p. 32-33).
- Black swans can be positive or negative, even if the latter are the ones which are more often discussed within the risk management context (p. 207).
- Professions which deal with the future and base their studies on the nonrepeatable past have an expert problem: experts need a static "tunnel", and dynamic environments are too prone to black swans to have real experts (p. 147).
- Predictions may be good at predicting the ordinary, but not the irregular, and this is where they ultimately fail (p. 149).
- Trying to predict specific black swans may make you more vulnerable to the ones you did not predict (p. 208).
- The concept of a grey swan which is a modellable extreme event while a black swan is a true unknown unknown (p. 272).
- In real life, we do not observe probability distributions. We just observe the events. Given a set of observations, plenty of statistical distributions can correspond to the exact same realizations (p. 353).
- There is no reliable way to compute small probabilities (p. 355).

The main problem with black swans is that classic risk management approaches could systematically ignore them due to their low probabilities. Aven (2013a) considers black swans as one of the basic features of new risk perspectives. Aven (2013d) argues that Taleb has a point when stressing the need for seeing beyond the standard probabilistic analysis when addressing risk, and that Taleb's book represent in this respect an important contribution despite lacking a proper scientific framing.

Makridakis & Taleb (2009) argue that there is always the chance of highly unlikely or totally unexpected occurrences materializing, and these can play a large role. Furthermore, except in artificial setups such as games, probability is not observable, and it is quite uncertain which probabilistic model to use. They also stress the difference between the *accuracy* of predictions and the *uncertainty* surrounding them - black swans introducing huge forecasting errors through the latter.

For Makridakis et al. (2009), black swans are a specific case of *coconut uncertainty*. They argue that psychologically, people are aware that rare events can occur and may even be able to imagine some examples, but they are unable to go far enough to consider their impact and consequences. They also make the important point that even though every particular black swan event is highly unlikely, this does not mean that the *class* of rare events that one typically fails to consider is negligible.

Feduzi & Runde (2014) define a black swan as "an unknown unknown that has gone on to occur", i.e. an event that the person who goes on to be surprised by it did not even imagine as a possibility prior to its occurrence. They ask the question, what is it about the world that gives rise to black swans, and propose two main sources: *emergence* and *epistemic constraints*. The former refers to emergent phenomena which cannot be reduced to a fixed set of prior causes and therefore foreseen even in principle (emergence is a key aspect of complex systems). The latter refers to the fact that the existence of surprises only requires limits on what the decision-maker can imagine: this is about the ability to collect and process evidence. Based on the distinction between these two possible sources, Feduzi & Runde propose that there are two different types of unknown unknowns:

- *Knowable* unknowns: these are related to epistemic constraints and therefore could at some point become known unknowns.
- *Unknowable* unknowns: these are related to emergent phenomena and cannot be transformed into known unknowns through more information.

Flage & Aven (2015) use the same (un)known-(un)known classification to clarify the difference between black swans and so-called *emerging risks*, defining the latter as *known* unknowns. Emerging risks are thus related to an activity where the background knowledge is weak but suggests that a new type of event could occur in the future and potentially have severe consequences to something humans value. They also highlight that *knowledge* becomes the key concept behind both black swans and emerging risks. And because knowledge can develop over time, time dynamics need to be considered in the context of these two concepts. Paté-Cornell (2012) makes a further distinction between black swans and *perfect storms*. Like black swans, perfect storms are extremely unlikely and can have a major impact. However, perfect storms involve mostly aleatory uncertainties (randomness) in conjunctions of rare but known events: unexpected combinations.

Higgins (2015) discusses black swans in the context of property asset management and notes that black swan events appear to be often overlooked by global organizations in their risk modeling frameworks when making major corporate property decisions. He also argues that such events are difficult if not impossible to model: "*It is even hard to imagine what kinds of events might fit into this category.*"

Sornette & Ouillon (2012) point out that many phenomena in the physical, natural, economic and social sciences are associated with power law statistics (with "fat tails"), and that this is also the domain of the black swans. For them, the black swan concept is pessimistic and "even dangerous as it promotes an attitude of irresponsibility", as these extreme events are considered unpredictable. They present another concept, the *dragon kings*. Dragon kings are defined as events which do not belong to the same population as the other events, in a specific quantitative and mechanistic sense. It is argued that dragon kings appear as a result of amplifying mechanisms that are not necessary fully active for the rest of the population. The key difference between black swans and dragon kings is that the latter could be at least partially predictable.

What are the main implications of black swans for risk assessment and risk management? Aven & Krohn (2014) argue that a broader concept of risk is needed to make risk management meaningful in a black swan world. Aven (2013d) considers essential to establish risk-uncertainty frameworks that are so broad that they also capture black swan events. Therefore, the knowledge dimension needs to be highlighted much more than what is typically seen currently in risk assessment applications. In this context, Aven also stresses the need to carry out a managerial review on top of the formal risk analysis, so that proper weight can be given to uncertainties and other concerns not captured by the formal assessment.

Aven (2015) discusses several approaches for managing black swans. First, two generic approaches are proposed: adaptive risk analysis and robust analysis. The former is about managing risks in constant interaction with the system, and using observations continually to adapt the risk management interventions. The latter is about building resilience into the system. Following the a)-b)-c) split of black swan events, Aven discusses specific considerations for the three types. For type a), focus on resilience and generating increased knowledge about the relevant phenomena is proposed. For type b), the issue is the limited knowledge of the analysts. Therefore, improved risk assessment and improved communication to relevant persons is proposed, so that both the analysis itself and the users of its results would be better informed. For type c), it is argued that it is appropriate to scrutinize both the judgments about acceptable risk and negligible probability, and the background knowledge that supports these judgments. In doing so, it should be acknowledged that:

- risk acceptability should not be based on probability alone
- even events with very low assigned probabilities may occur
- cautionary and precautionary principles should be applied.

The first bullet is a fundamentally important consideration as it *forces black swan events outside the classic risk management approach*. Paté-Cornell (2012) agrees by arguing that black swan -type of risks need to be explicitly treated through risk analysis. They should not be ignored by saying that they are too unlikely to be accounted for. However, she points out that it is challenging to represent accurately, in the risk results, the existing information about epistemic uncertainties, especially when there are disagreements among experts. Concerning the third bullet, the cautionary principle states that in the face of uncertainty caution should be the ruling principle, and the precautionary principle may be considered a special case of the cautionary principle, applying to scientific uncertainties (Aven 2015).

Aven (2015) points out that the classic approach of using probability limits for acceptability of risks is not in general justified as it ignores the degree of knowledge that supports the probability assignments. This is another fundamental point, because as a consequence, risks cannot simply be placed in a probability-severity-space where risk acceptability could be a simple line. Judgment of acceptability of risks needs to become something more refined and multidimensional. *Uncertainty analysis* can also be a supporting factor, as discussed in Chapter 2.2.

4 Accommodating different risk aversion policies

Sometimes one factor which influences the tolerability of a risk is whether the potential loss is a one-time big loss or several smaller losses paced in time. This leads to the topic of risk aversion. The foundation of risk aversion is in *loss aversion* which Rabin & Thaler (2001) describe as the tendency to feel the pain of a loss more acutely than the pleasure of an equal sized gain. Loss aversion is part of the prospect theory by Kahneman & Tversky (1979), modeling decision-makers who react to changes in wealth, and are roughly twice as sensitive to perceived losses than to gains. To quote Kahneman & Tversky: “the aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount.” Kahneman & Tversky (1979) argue that even if their prospect theory concerned mainly monetary outcomes, the theory should be readily applicable to choices involving other attributes, for example quality-of-life or the number of lives that could be lost or saved related to a policy decision.

Aven (2011a) is critical of the definition for risk aversion offered by the International Organization for Standardization: “attitude to turn away from risk” (ISO 2009). He argues the meaning of the definition and in particular the expression “to turn away from” is not clear. Aven points out that there is already a common definition for risk aversion used in risk analysis and in economic and business applications: risk aversion means that a decision-maker’s *certainty equivalent* is less than the expected value. Here the certainty equivalent refers to the amount of payoff that needs to be obtained in order to be indifferent between that payoff and the actual gamble.

More specifically, in risk management the risk aversion concept is reflected in the question on whether experiencing multiple fatalities in a single event is more serious than if the same number of fatalities are distributed over several events (Johansen & Rausand 2014). Decisions need to be made about what kind of policy in relation to risk aversion is adopted. If a neutral policy is adopted, a single high-fatality accident is not considered more important than several smaller accidents causing the same total number of fatalities.

Rheinberger (2010) reports the popular tendency to give additional weight to large consequences and argues that this practice is based on the assumption that the costs of large accidents to the society are never fully reflected by the expected direct losses - therefore risk aversion acts as a measure of precaution. However, he challenges this practice, arguing that risk-averse behavior may be neither appropriate for managing multiple fatality accidents, nor necessarily of benefit to the public. His own results show that neither experts nor lay people display risk-averse preferences in comparative assessments of mortality risks. Indeed, in his experiments, sometimes more frequent accidents involving

one or two fatalities were judged to be far less acceptable than less frequent accidents involving 4 to 6 fatalities.

The main reason for discussing risk aversion here is that choosing a specific risk aversion policy has implications on risk management. For example, if the priority is on avoiding large consequences, then there needs to be a mechanism which gives more weight to these threats/scenarios/events. If the policy may vary, then the method needs to be able to accommodate various risk aversion policies.

From another perspective, being risk-averse can be paired with the corresponding opposite concept of being *risk seeking*. These are both different *risk attitudes* and another way to express this is to talk about high or low *risk appetite* - even if the latter term is less properly defined (Aven 2013c). A high risk appetite can be interpreted as a willingness to take big risks in pursuit of values. All these concepts have strong links to different human biases and tendencies in terms of understanding and assessing different risks and opportunities.

5 Synthesis and conclusions

For a long time, risk has been understood as a product of probability and severity. This view of risk still dominates a large part of practical applications within industries. However, a new modern view is in the process of replacing the classic understanding of risk. The new risk perspectives emphasize uncertainty and the knowledge dimension – and complement the probability-based approaches by integrating the strength of knowledge and the potential suprises into the risk assessment. This modern view of risk together with the new risk perspectives are adopted in this dissertation. In line with that, probability is understood as a *subjective probability*, with a reference to an uncertainty standard.

Uncertainty was discussed in length above, and it was noted that uncertainty analysis can be used to assess the level of uncertainty in any particular situation. Classifications for uncertainty levels exist. In the context of risk assessment, it was argued that the strength of knowledge concept reflects more precisely the right ideas than the uncertainty concept itself. The important lessons within the new approach to risk assessment/management are that there needs to be a strong focus on knowledge building in order to minimize the uncertainty in risk assessment; reliance on expert judgment is necessary; all assumptions need to be described clearly;strength of knowledge needs to be assessed; results will be comparable only to the extent that they have been computed on comparable bases; (strength of) knowledge becomes a real dimension of risk and will have an impact on the result of the risk assessment; surprises need to be addressed as a part of the risk assessment/management process. Specifying how the uncertainty level will be integrated with the results of the risk assessment is an important question in its own right.

Already due to these new requirements, it becomes clear that the result of the risk assessment cannot be a simple number. A related point is the statement by Aven (2015) that the use of probability limits for acceptability of risks is not in general justified. Aven (2013d) also argues that risk assessment as such is often not enough for decision-making: what he calls a *managerial review* is also needed in order to take into account all the different relevant aspects. All this together means that presenting the results of risk assessment and taking decisions based on the results, becomes something more complicated and multidimensional than just comparing two numbers. Yet another reason for avoiding simple numbers to describe risk levels is the question related to risk aversion: how could a number reflect the chosen risk aversion policy? For example, high severity events might have to be prioritized over several lower-severity events. It is argued that there is value in trying to present the results in such a way that the risk aversion policy is left open and can be adjusted according to various needs – especially that as shown above, there is no universal agreement on what the risk aversion policy should be.

Surprises are a key part of the new risk perspective and are mainly represented through the concept of black swan. It was noted that even though every particular black swan event may be highly unlikely, it is not very unlikely that *one* of the potential black swans materializes. There are slight variations in how

a black swan is defined: Taleb includes *unknown unknowns* and unknown knowns whereas Aven opens the definition to include risks that were on the radar but were dismissed due to their very low probability. Feduzi & Runde (2014) define *unknown unknowns* as subjective to the decision-maker, which means that *unknown knowns* are included as *knowable unknowns*. In this dissertation, the a)-b)-c) list of Aven (2015) is used as the main reference. For the practical purposes related to the methodology presented in Part II, these slight differences do not pose any problems. Low probability - high impact risks would sit together in a specific area of the risk picture, including all the types that Aven lists. Strictly speaking, it could be argued that only the *unknown unknowns* and *unknown knowns* would be considered true black swans. The latter are related to the *local* lack of knowledge and could also be called knowable unknowns. Aven (2013b) also suggests that in the presence of strong knowledge, there should be no black swans. It could be argued, that this is really a question of what is meant by strong knowledge: in a complex environment it is difficult to imagine that knowledge would be so vast as to completely rule out the possibility of black swans. In fact, it could even be argued that a *perception* of strong knowledge with a few gaps could be a recipe for major surprises - this is also what Taleb suggests through his point on *epistemic arrogance*.

Perfect storms and *dragon kings* were also mentioned. Due to their nature, they are defined in the context of systems where modeling is at least somewhat possible, e.g. in the context of physical phenomena governed by deterministic laws of nature, or at least within the reach of some statistical methods. Therefore, it is argued that the black swan concept is particularly appropriate in the context of complex systems where the phenomena emerge within a *system of sociotechnical systems*, and typically escape any available statistical methods. It is clear, that the proposed risk management framework will have to be able to maintain awareness of potential black swans, irrespective of their perceived probabilities. This may be a challenging requirement because as mentioned above, classic risk management approaches could dismiss potential black swan events due to their extremely low probabilities.

Identifying and endorsing all the key aspects of the modern risk perspective does not mean that every real-life risk assessment can be carried out respecting all the new requirements exhaustively. Real life is characterized by constraints, e.g. on resources, and this means that intelligent compromises may have to be applied to define methods and processes which are acceptable both from the scientific and pragmatic points of view.

The uncertainty in risk assessment links strongly with the topic of the next chapter: complexity. It is easy to argue that the (frequentist) notion of repeatable similar experiments is very far from the reality of a complex transport operation which is under constant evolution in many different dimensions. Complexity also means that many of the unknowns are *unknowable unknowns*. Deep uncertainties mushroom. This situation means that even experts cannot advise on many key questions.

Chapter 3 – Risk Management in a Complex Adaptive System

This chapter lays the foundations for treating the transport system as a Complex Adaptive System. Complexity has two main implications on risk management. First, for risk assessment, complex systems are particularly difficult because uncertainties and the potential for surprises increase. Secondly, for risk treatment, if one wants to achieve positive change in a complex system, the ways to intervene must be adapted to the complexity.

The chapter develops the modern framework on complex adaptive systems step-by-step, starting with an overview of the Cartesian reductionism in Chapter 3.1 and then developing the properties of complex systems in Chapter 3.2. and specifics of complex human systems in Chapter 3.3. The contribution of

systems thinking to understanding complex systems is discussed in Chapter 3.4 and finally Chapter 3.5 draws the conclusions in terms of what could be the best strategies for dealing with complex systems.

Russell Ackoff, an honored contributor to the development of *systems thinking*, defined a *system* in the following way:

> "A system is a set of interrelated elements. Thus a system is an entity which is composed of at least two elements and a relation that holds between each of its elements and at least one other element in the set. Each of a system's elements is connected to every other element, directly or indirectly. Furthermore, no subset of elements is unrelated to any other subset" (Ackoff 1971).

Furthermore, he defined:

- The *environment* of the system as a set of elements and their relevant properties, which elements are not part of the system but a change in any of which can produce a change in the state of the system.
- A *closed system* as one that has no environment. An open system is one that does.
- A *purposeful system* as one which can produce the same outcome in different ways in the same (internal or external) state and can produce different outcomes in the same and different states. Thus a purposeful system is one which can change its goals under constant conditions; it selects ends as well as means and thus displays will.
- An *organization* is a purposeful system that contains at least two purposeful elements which have a common purpose relative to which the system has a functional division of labor; its functionally distinct subsets can respond to each other's behavior through observation or communication; and at least one subset has a system control-function.

Another renowned researcher in the field of systems thinking, Donella H. Meadows, gives the following definition:

> "A system is an interconnected set of elements that is coherently organized in a way that achieves something" (Meadows 2008).

She points out that a system must consist of three kinds of things: *elements*, *interconnections*, and a *function* or *purpose*, where the purpose is not necessarily spoken, written, or expressed explicitly: "purposes are deduced from behavior, not from rhetoric or stated goals". Furthermore:

> "System purposes need not be human purposes and are not necessarily those intended by any single actor within the system. In fact, one of the most frustrating aspects of systems is that the purposes of sub-units may add up to an overall behavior that no one wants" (Meadows 2008).

The *elements* of the transport system would typically include cars, private people, aircraft, professional pilots, roads, agencies, computer systems, passengers, trains, trucks, taxi drivers, truck drivers, ships, crewmembers with their dedicated roles, politicians dealing with transport matters, and so on. The *interconnections* would include all the communications between individuals and organizations, the laws and regulations governing the activity, operating procedures, policies, marketing, and even laws of physics acting on vehicles, ships and aircraft, and so on.

Scientific disciplines such as system dynamics, systems thinking and operations research have greatly enhanced the understanding and improvement of systems (Forrester 1994). They have focused on systems as such and shown that systems have properties which are only apparent and meaningful at the level of the whole and cannot be induced from the parts.

The understanding of systems and especially complex systems has greatly increased in recent times. Just as the new risk perspectives are becoming better known, new thinking on complex systems has challenged the very strong tradition of Cartesian- Newtonian thinking.

1 From Cartesian reductionism to complexity

The heart of the Cartesian-Newtonian understanding of a system is that any system can be reduced to its parts and all system properties are fully defined by the parts. Such a reductionist methodology is strictly analytical: everything about the whole, including how it functions, can be learned and explained by finding the parts (Gilbert & Sarkar 2000).

Reductionism can be associated with the so-called machine metaphor: a machine can indeed be reduced to its parts without losing its machine-like character (Mikulecky 2001). If the components under study didn't explain the behavior of the machine, the reductionist approach would be to split the components further into even smaller parts, and continue this way until the whole can be understood (Dekker 2011, p. 57).

As the names Newton and Descartes (behind the term *Cartesian*) indicate, reductionism and the whole world view that goes with it has its origins in the scientific revolution. Indeed, impressive progress was made in science during that time, and since that time, using the reductionist approach. This success in explaining natural phenomena hinted that *nature as well was a machine* and could be explained by observing its parts. There was an idea that if science was applied correctly it would lead to absolute certainty (Dekker 2011, p. 56).

According to Dekker et al. (2011), the Cartesian-Newtonian thinking also includes the following ideas:

- Everything that happens has a definitive and identifiable cause, and a definite effect.
- There is symmetry between cause and effect: a significant effect can only be produced by a significant cause.
- Knowing the current state of the system in detail and the laws governing it, makes it possible to predict its future state with absolute certainty.
- The previous point also means that people within the system should be aware what the future brings, i.e. if something negative happens, someone must have seen it coming.
- On the other hand, one can also reconstruct the past state of the system based on its current state.
- In other words, time is reversible.

These points draw a picture of a universe which is predictable (Gershenson 2013). Mikulecky (2001) argues that the so-called *hard science* and its *scientific method* are associated with the description of Cartesian reductionism and the machine metaphor. He argues that the machine metaphor set the tone for modern science and has lasted since that time. Gershenson (2013) agrees and writes that this classical scientific worldview still dominates science and philosophy. He refers to textbook exercises where systems are usually closed, ideal conditions are assumed, and elements are "well behaved".

Why is the effort made here to discuss reductionism in the context of complexity? The world-view based on reductionism is much more intuitive than complexity thinking, and it is very much present in today's thinking. As such, it presents a major hurdle for embracing complexity and adopting the associated methods.

There are several examples where people point out the existence of reductionism in today's world. Beckerman (2000) argues that "systems engineering, as it is practiced today, is almost exclusively reductionist". She points out that the consequences of the reductionist approach go through the entire development lifecycle and gives examples on system architecture and system specifications. She writes: "the sum of the behavior of the individual components is assumed to provide the system properties. An assumption or approximation of linearity replaces nonlinearities. Interactions between components are assumed to be few, weak, and linear".

Dekker et al. (2011) discusses the application of cartesian linear thinking in investigation of failures in complex systems. He gives examples of single factor, judgmental explanations for complex system

failures made in aviation, medicine, military operations and road traffic. According to Dekker, the explanations of such accidents often use language such as *chain of events* and *human error* and questions such as "what was the cause". One of the consequences of the Newtonian approach is that there can be *only one true story* of what happened (Dekker 2011, p. 83). Dekker also points out that the very idea of a *root cause* is Newtonian. It should be obvious that if problems in the complex world are approached with a simplistic Newtonian view, the results will not be good. Risks in a complex system should be treated based on an understanding of complexity.

Both space shuttle accidents can serve as examples (CAIB 2003, Reason 1990). The Newtonian approach can easily determine the technical causes of the accidents (the O-rings for Challenger and the detaching insulation foam for Columbia) but the real interesting questions emerge when one recognizes that both technical conditions had been identified *years* before the accidents, documented and treated in line with NASA's processes. It is obvious that NASA did not lack technical expertise. The problems had been detected repeatedly and their high-risk nature recognized. Locally, each decision, taken one at a time, seemed correct. The accidents were not preceded by a "smoking gun" or a perceived dangerous "component" but by "business as usual". Because of all this, from a Newtonian perspective, these accidents should not have occurred. But they did. The CAIB report tells a long story involving NASA's budgets, history, program culture, politics, compromises, changing priorities of the democratic process and an intense pressure on the program to stay on schedule. NASA was (and is) part of an open system, involving for example the Congress and the White House. It is not possible to understand the pressures experienced inside NASA without looking "up and out" (synthesis) from NASA, rather than only focusing "down and in" (analysis), to use Dekker's (2011, p. 130) terminology. The shuttle itself was already a compromise which never met its original requirements for reliability, cost, ease of turnaround, maintainability or safety. Political, budgetary and policy decisions from above impacted the program's structure, culture and safety systems and resulted in flawed decision-making related to both accidents (CAIB 2003, p. 195). Both accidents featured "normalization" of the detected anomalies, so flying with the flaws became routine and acceptable. It was pointed out that NASA did not have a truly independent safety program and did not demonstrate the characteristics of a learning organization. It is noteworthy that NASA had been unable to change even after the Challenger accident.

Ruhl (1996) discusses the relationship of complexity with the law and points out that:" reductionist approaches to legal administration have produced a system that, in the lexicon of nonlinear dynamical systems theory, is sitting predominantly on the non-adaptive, static regulations attractor in a futile effort to increase the predictability of the law-and-society system. This has been accomplished at the expense of freedoms and rights, which are more adaptive in nature than are regulations". In other words, reductionist approaches also produce rigid, inflexible regulations which do not evolve quickly enough when the system itself evolves. Regulation is also an essential part of safe transport systems like aviation. Therefore, the lesson of adaptability is very relevant, and this necessitates adopting the complex world view.

As Gershenson (2013) points out, recent advancements in science have made a reductionist worldview obsolete. Why is this old world-view still so popular? Dekker et al. (2011) point out that such thinking has long been equated with science and rationality in the West and states: "the mechanistic paradigm is compelling in its simplicity, coherence and apparent completeness and largely consistent with intuition and common sense". The Cartesian-Newtonian assumptions remain largely transparent and unchallenged in safety work precisely because they are so self-evident and common-sensical (Dekker 2011, p. 84). Inayatullah (2002) argues that there are clear market demands for reductionism. According to him, students, governments and business organizations tend to desire a *single future* giving them a clear answer. Adopting an understanding of complexity would require accepting that there are many factors explaining change and that there will always be some unknown factors. There are thus psychological barriers to letting go of the ordered, Newtonian world-view.

Mikulecky (2001) points out that most of modern science and technology is the result of the success of the Newtonian paradigm, and therefore it cannot be ignored. He argues that the complexity perspective

is the result of the failure of the Newtonian paradigm to be generic. The worldview of simple mechanisms, for him, is a fictitious world created by the hard version of science as a formal system that hopes to model the real world. Real-world complexity does not fit in the reductionist worldview. In complex systems, there are system effects which cannot be derived from the parts (Urry 2010). In a complex system, *interactions* between the components are fundamentally important. Properties that are not those of any of the parts, but which arise through the interactions are called *emergent properties* (Gilbert & Sarkar 2000).

Safety culture is a good example of an emergent property. A specific kind of safety culture cannot be implemented by force or by bringing in certain components which would somehow create the desired culture in any given organization.

Before discussing the properties of complex systems in detail in the following section, it is important to make the difference between systems which are only *complicated* compared to truly complex systems. A complicated system is not easy to understand because it is not simple. Typically, a complicated system consists of a huge number of parts, an example could be a modern passenger aircraft. Due to the amount of information and detail, such a system cannot practically be understood completely by a single person. However, a complicated system is *understandable and describable in principle*, and it can be taken apart and put together again (Dekker et al. 2011). In other words, each part has its function and the behavior of each part is fully predictable, and so is the behavior of the whole system. A complicated system is in line with the machine metaphor and reductionism works perfectly. In contrast, as Mikulecky (2001) states, complexity is the property of a system that is manifest in the inability of any one formalism being adequate to capture all its properties.

2 Properties of complex systems

There are many overlapping descriptions of complex systems. Here, the numbered list of 10 characteristics of complex systems from Cilliers (1998) can be used as a good first overall definition of a complex system:

> (i) Complex systems consist of a large number of elements.
>
> (ii) Having a lot of elements is a necessary but not sufficient condition for having a complex system. The elements have to interact dynamically.
>
> (iii) The interaction is fairly rich, i.e. any element in the system influences,
> and is influenced by, quite a few other ones. Not all the elements in the system, however, need to possess interactions with a large number of other elements - this can vary from one element to another.
>
> (iv) A lot of the interactions are nonlinear. This means, that small inputs in one part of the system can have large results in other parts of the system and vice versa.
>
> (v) Each element is mainly dealing with its immediate neighbors in its local context. However, due to the large number of interactions between elements, influences can propagate easily from one part of the system to other parts of the system. Thanks to the previous point, it is clear that influences get modulated along the way: e.g. they may be enhanced, suppressed or altered in many different ways.
>
> (vi) A fundamental aspect of a complex system is that it has *feedback loops*. Any effect can feedback onto itself through one of several feedback loops. The feedback can be positive (enhancing, stimulating) or negative (inhibiting, dampening).

(vii) Complex systems are usually open systems. This means they have live interactions with the environment and there are a multitude of influences between the system and its environment. In fact, typically the system cannot be truly extracted from its environment: any definition of the system borders is just a theoretical convention. The system borders could be defined in many ways depending on the purpose and the position of the observer. In contrast, closed systems are typically complicated, not complex.

(viii) Complex systems operate under conditions far from equilibrium. The system is highly dynamic and constantly evolving.

(ix) Complex systems have a history. Not only do they evolve through time, but their past is co-responsible for their present behavior.

(x) Each element in the system is ignorant of the behavior of the system as a whole, it responds only to information that is available to it locally.

Mikulecky (2001) examines systems in the light of Aristotle's four types of causes: *material*, *efficient*, *formal* and *final* causes. He argues that in the context of complex systems the *final cause* ("why") becomes the key factor for understanding the system:

> "In the complex world, function becomes the *dominant* aspect of the system's essence and the material parts are suppressed to a much less important role."

This is in line with what Meadows stated above about the purpose of a system. To achieve its purpose, the system adapts continuously. This is the basis for the concept of a *Complex Adaptive System* (CAS). In most practical cases of interest, the complex adaptive system would be an organization or a system of several organizations. Reiman et al. (2015) present the following key features of complex adaptive organizations:

- *Non-linearity*: Inputs are not necessarily proportional to outputs. Systems are composed of highly responsive and interconnected feedback loops that can reinforce or attenuate inputs. All effects have several parallel contributing factors, instead of one or few causal chains as in linear systems. There are 'spiraling, iterative cycles of cause and effect' instead of one root cause for each effect. On the other hand, complex adaptive systems also exhibit time delays between 'causes' and 'effects', which can lead to overshoots in interventions

- *Emergence*: As a consequence of the interactions among diverse agents, new patterns of relationships, new system level properties and structures emerge. Emergent properties forming from the interaction of the agents cannot be traced back to those individual agents. Yet these patterns have an effect on the agents. The irreducible nature of emergent properties means that the properties of the whole are distinctly different from the properties of the parts.

- *Self-organization*. Self-organization denotes the emergence of new structures, patterns and new forms of behavior in the system as a consequence of agent interaction and connections. Organizations are continually self-organizing through the processes of emergence and feedback. Thus, the phenomenon of self-organization is the collective (emergent and ever non-permanent) result of local yet non-linear interactions among agents. Complex adaptive systems can thus self-organize into even greater states of complexity.

- *Far-from-equilibrium conditions*. This is sometimes called the edge of chaos, the condition of high requisite variety and creativity. Being far from equilibrium also means that the system is in a continuous process of flux and change. Change in these systems is a natural tendency, not something initiated by an outside force. This capability also allows these systems to self-organize and adapt to changes in their environment.

- *Coevolution*. A complex adaptive system exists within its environment, but it is also part of that environment. Environmental changes require a change in the system. However, since the system is part of its environment, change in a system changes its environment, creating a process of mutual change and evolution. Further, the environment including the organization can be considered a CAS of its own, which also learns and adapts (see nested systems).

- *Nested systems*. Complex adaptive systems are sometimes called 'systems within systems'. The nested systems increase the diversity and uncertainty inherent in the 'parent system'.

- *History-dependence*. A CAS cannot be rewound back to its earlier form and state. Actions are thus irreversible, and the past helps to shape present behavior. Agents learn from their previous experiences and change their actions accordingly. History dependence also means that solutions can seldom be copied from one system to another: what works in one organization cannot be replicated in another organization, since they each have their own distinct histories.

The term "agent" used by Reiman et al. corresponds to "element" used by Cilliers. Reiman et al. (2015) describe a Complex Adaptive System in the following way:

> "A complex adaptive system is a collection of individual *agents* with freedom to act in ways that are not always predictable, and whose actions are interconnected so that one agent's actions change the context for other agents. These agents interact in a non-linear way creating system-wide patterns and higher and higher levels of complexity. The agents differ from each other and none understands the system in its entirety. This diversity is a source of invention and improvisation. As the agents are interdependent of each other, relationships among agents can be considered to be the essence of a complex adaptive system. Understanding a complex adaptive system requires understanding of patterns of relationships among agents."

Depending on the perspective, various capabilities of complex adaptive systems can be highlighted as the fundamental defining characteristics of these systems. Perhaps the most important aspect in this respect is the phenomenon of *emergence*. Emergent properties are properties at the level of the whole which cannot be predicted on the basis of the components. It is often stated that a complex system is more than the sum of its parts (e.g. Cilliers 1998, p. 2). Urry (2005) rephrases this by stating that there are *system effects* that are different from their parts. Consequently, like Mikulecky (2001) states, "the essence of complexity is in the existence of something that is lost as the system is reduced to its parts". This also means that the machine metaphor cannot be applied to a complex system. Typical examples of emergent phenomena include consciousness emerging in the brain (which consists of simple neurons) and climate in a work group (Reiman et al. 2015). Emergence, especially together with the inherent nonlinearity of complex systems, often creates unexpected effects, surprises. Urry (2005) also mentions *tipping points* as an additional mechanism through which unexpected structures and events may emerge.

Another fundamental characteristic of a complex adaptive system is the existence of *feedback loops* and their impact on *causality*. The feedback loops introduce a circular element to causality, which combines with the nonlinearity and emergence. As a consequence, there is great *dependency* between the elements within a complex system but cause and effect in their classic sense lose their essence (Le Coze 2005).

Cilliers (1998) argues that due to the necessity of dealing successfully with a changing environment, there are two capabilities which are indispensable for complex systems: first they must be able to *store information* for future use, and secondly, they must be able to *adapt* their structure when necessary. The first capability can be called the process of *representation* and the second the process of *self-organization*. Mikulecky (2001) states that the system has within itself a model of its environment and uses this model to influence *present* behavior in *anticipation* of future events.

The capacity to learn and anticipate also links with another characteristic of complex systems: *path dependency* (history-dependence in the above list). The system learns from its past experiences and this will have an impact on its behavior in the future. Like Cilliers (1998) states: "no complex system, whether biological or social, can be understood without considering its history". He argues that the history is important not only for understanding the system, but it also has an impact on its *structure*. Path dependency (together with feedback loops and emergence) leads to *time-irreversibility*. Ordering of events through time influences significantly the non-linear ways in which they eventually turn out (Urry 2005). This also means that it is impossible to reconstruct the past, i.e. to establish the precise set of conditions that gave rise to a specific outcome (Dekker et al. 2011).

As stated in the definition of a complex adaptive system by Reiman et al. (2015), none of the agents in the system understand the system in its entirety. In other words, information and interactions in the system are local and it is impossible for anyone to understand the system fully. As stated by Dekker et al. (2011), both the knowledge of initial conditions and the knowledge of the laws governing the system are unobtainable in complex systems. Mikulecky (2001) argues that no one formalism is adequate to capture all the properties of a complex system. Indeed, the only hope in trying to understand a complex system is to try to gather as many *different perspectives* as possible, e.g. through multiple narratives (Dekker et al. 2011).

Thanks to all the mentioned characteristics, complex adaptive systems *behave in unpredictable ways*. This unpredictability is a fundamental characteristic of complex adaptive systems. It would be fallacious to think that some new method could reveal the whole complexity and enable prediction of the functioning of a complex adaptive system (Reiman et al. 2015). Any kind of *analysis* of a complex system will always distort the reality by cutting out parts of the system. The only reliable model of a complex system would be the system itself (Cilliers, 1998, p. 24).

3 Organizations as complex adaptive systems

Organizations are in a central role in the transport system. Operators are organizations. Agencies are organizations. Safety and risks emerge within organizations and organizations can be a promising level of treating many risks efficiently if there are ways to increase organizational resilience. Therefore, it is important to understand organizations as complex systems and organizations within a complex system.

Le Coze (2005) presents several reasons why organizations cannot be effectively analyzed through their parts in a reductionist manner, for example:

- Organizations are not closed systems. They are open and adaptive, they evolve and drift.
- Organizations don't follow linear cause effect relationships
- They have complex interactions between their parts
- They are difficult to predict through quantitative models. The number of interactions and autonomous variables is too high.

In a similar manner, Cilliers (1998), uses his description of a complex system to illustrate how people in a country (in this case taken as economic agents) form a complex system (p. 6) and how a society is also a complex system (p. 119). Using the same criteria, it could also be shown that an airline, a national safety agency, the transport ministry, or the national transport system (consisting of these and many other agents) are all examples of complex systems. Such complex systems of humans, or simply organizations, have some specific characteristics not yet discussed in the previous section, yet significant for the topics of risk management and safety.

The way things happen in an organization is harder to observe or trace retrospectively, because the interactions between human agents are not as regular and concrete as tangible artifacts. Dekker et al. (2011) argue that decisions, for example, are influenced by the "messy" organizational life which consists of not only written protocols but also unwritten routines, implicit expectations, professional judgments and subtle oral influences. Le Coze (2005) also refers to interactions, rituals and myths of the

social structure, and beliefs of the entire organization. He also states that organizations cannot be totally described through plans and organizational charts: the real complexity cannot be captured "through arrows and boxes drawings". When La Porte (2006) lists the conditions for effective adaptive response, besides the quite tangible point of having sufficient resources, he also mentions more intangible and invisible factors: *trust* among leaders, organizations and other people; as well as *sufficient support* from the community at the interface between the organization and its environment. He goes on to point out that these conditions can be measured technically, *organizationally*, and *culturally*. Snowden (2005a) states that "it is not infrequent to discover that an informal community, such as an officer's mess or a football supporters club, provides a critical mechanism for learning". These examples illustrate the difficulty of capturing all the relevant factors influencing the way an organization evolves, adapts and interacts with its environment, and the various scientific disciplines that would need to be considered in order to ensure having a sufficient number of different perspectives.

Cilliers (1998, p. 94) argues that besides *cooperation* there is also *competition* within an organization. According to him, competing for limited resources is the basic driving force behind the development of structure. Dekker & Nyce (2014) make the point that life in an organization is not always harmonious - *conflict* is a natural part of organizational life. This brings them to discuss the issue of *power* within the organizational context, and especially in relation to safety. They criticize safety science for taking a very idealistic view to power: the burden is placed on low ranking people who are expected to speak up and "talk truth to power". Organizational failure can even be allocated to a few individuals who "didn't speak up ". Senior management is assumed to be committed to safety more or less because that should be their role. Dekker and Nyce argue that what really counts is where the power is: distribution of scarce resources, deciding what is on the agenda and who has the power to win internal competitions. They argue that trying to go against the power structures through training does not remove the problem. Part of the power is to decide what is the official narrative when something has gone wrong - assigning cause is often a political construct and power play leads to scapegoating. Making a specific risk recognized also requires power. According to Dekker & Nyce, "power is everywhere" and using safety culture to define organizational values may have a negative effect on the diversity by silencing conflicts and contradictions, and often forcing organizations into a very narrow-minded normative view of organizational (safety) culture. Antonsen (2009) agrees that the role of power in organizations is rarely addressed in safety culture research.

The way the complex system behaves at the system-level has a lot to do with the way in which its agents interact. This again, is influenced by factors at the level of the agents. In the case of organizations, the agents are often human beings, and as such subject to all the biological, psychological, etc. characteristics and limitations of humans. The invisible nature of many of these factors and power - discussed in the two preceding paragraphs - are just examples of the many factors in play. Anderson (1999) mentions some of the factors related to agents:

- Each agent adapts to its environment. This makes agents *co-evolve* with one another.
- The *linkages* between agents may evolve over time. The strength, functional form and the pattern of interconnections may thus change.
- Agents may *leave* the system, new agents *enter*, and some agents *transform*. These changes make the complex adaptive system evolve.

Marion & Uhl-Bien (2001) examine complex organizations from the perspective of *leadership*. They argue that existing approaches to the study of leadership have been heavily grounded in the premise that leadership is interpersonal influence and have therefore focused primarily on leader attributes and emotions of the followers. They promote the view that complexity science broadens conceptualizations of leadership to include processes for managing dynamic systems and interconnectivity. Their research is reviewed here because it is seen to have value both for understanding the behavior of complex systems and for influencing them successfully. Marion & Uhl-Bien use the term *complex leadership* and associate it with:

- The idea that leaders *cannot* control the future with deliberate interventions because in complex systems the dynamics which determine future conditions are complex and unpredictable.

- Creating the conditions that enable productive, but largely unspecified, future states; as well as emergence of *distributed intelligence*.
- Influencing networks and creating atmospheres which catalyze the formation of aggregates and meta-aggregates and emergence of innovations and their dissemination.
- Cultivating largely undirected interactions among individuals and work groups to enable common understanding and the creation of favorable uncontrolled futures and productive surprises.
- Understanding organizational behavior in terms of *global interactions* rather than narrow focus on controlling *local events*.
- Enabling effective *networks* rather than trying to *motivate* enhanced effort or keep peace.
- Using interaction to enable individuals and work groups to work through *conflicting constraints.*
- Learning to manipulate the *situations of complexity* rather than the results.

Marion & Uhl-Bien argue that emergent structures are produced by a combination of (bottom-up) microdynamic and macrodynamic forces. When individuals interact, both coordinated and random behavior occurs. These interactions create linkages which may evolve into an aggregate (i.e. a combination of linkages) which can be considered a system in itself. For example, a family unit or a work group is an aggregate. The actors within share a sense of common identity and need to be able to resolve conflicting constraints. The latter requirement means there is a practical size limitation to an aggregate as too many conflicting constraints would paralyze the aggregate. The aggregates may create meta-aggregates which in turn may create meta-meta-aggregates and so on. Any of these can be called ensembles. The *macro level* is about the dynamics between the ensembles. This behavior is argued to be self-generative, bottom-up and non-linear. Persistently interacting social networks create order, innovation and fitness, but they elude control and prediction. Aggregates within a meta-aggregate are linked together through direct dependence on common resources or events. Typically, the average *coupling* between aggregates within a complex system is *moderate*: too strong coupling would impose too many conflicting constraints. Marion & Uhl-Bien argue that *command and control leadership* is a barrier rather than a success factor in the context of complex systems, and that this applies even to charismatic leadership. However, actively involved leadership can be a positive factor if it's done right. Importantly, it is argued that innovation and fitness are better served by bottom-up rather than top-down coordination. At the practical level, complex leadership can consist of delegation, encouragement, resources, avoiding interfering personally, using symbols (or "tags"), encouraging people to try things, to evaluate and change their experiments. Knowledge centers within the organization should be identified and encouraged to do creative things and communicate with one another. Ideas and creative surprises can also be sought from outside, e.g. by attending conferences.

Snowden (2005b) makes a difference between *ordered* and *un-ordered* systems, complex systems being part of the latter. He argues that management science and consultancy practice are dominated by approaches based on an assumption that the system under study is ordered in nature. According to him, this has led to the application of methods based on best practice and structured top-down approaches but trying to apply such approaches to un-ordered systems "involves deploying major resources to make things worse". He argues that unordered systems "work bottom-up not top down" and that "patterns emerge from the interaction of many agents operating on unarticulated rules with other agents and with the environment". La Porte (2006) makes the following sharp observation:

> "Literature on highly reliable organizations does not cover the way in which organizations come to possess their special characteristics, or what managers have done, not to mention what they should do, to steer their organizations in this direction."

Taking the points both from Marion & Uhl-Bien (2001) and Snowden (2005b), the proposed answer is that such organizational properties cannot be constructed solely by direct top-down command. The next step in reasoning is the following fundamentally important statement (Reiman et al. 2015):

> "Safety is an emergent property of the complex adaptive organization and as such it cannot be standardized or controlled."

The same statement could of course have been made of any emergent property of a complex system, but in the current context of risk, this fundamental observation of safety in a complex adaptive system is highly relevant and has significant consequences. Furthermore, Reiman et al. (2015) point out that *control* in complex adaptive systems is always *distributed* rather than centralized. Consequently, management has to focus on creating *preconditions* and *organizational potential* for safety instead of trying to control or command employees. It can be seen, that Marion & Uhl-Bien (2001), Reiman et al. (2015) and Snowden (2005b) are in agreement on the role of control and management within complex adaptive systems.

Not surprisingly, emergence can also have negative implications. In the context of safety and risk, at least two phenomena are worth mentioning. The first one is called *normalization of deviance*. It refers to a process where small deviations from the standard practice gradually become the accepted norm (Reiman et al. 2015). The key point in this process is that the valuation of the deviance changes from unacceptable to acceptable. This also means that potential danger signals are neutralized because they have become part of normal practice. The second, closely related, phenomenon is *organizational drift*. Organizational drift is about local adaptations and modifications to the centrally designed standard practice. Sidney Dekker dedicated a full book *Drift into Failure* (2011) to the combined impact of these two emergent phenomena. Reiman et al. (2015) list several factors which can contribute to drift and its normalization:

- Tight resources, and the resulting pressure to optimize
- (Real or perceived) production pressures
- Uncertain and unruly technology (e.g. false alarms)
- Structural secrecy (danger signals remain local)
- Intolerance of dissenting opinions
- Distant information patterns (information in the system is not widely understandable)
- Loose couplings within the system make local adaptations possible

The two space shuttle accidents cited above offer good examples of normalization of deviance, as well as Naweed et al. (2015) in the context of railway operations safety.

The ultimate level of failure, at least from the safety standpoint, is the *system accident*, i.e. an accident which is largely caused by the interactive complexity of the system itself. In his visionary book, *Normal accidents* (1984), Charles Perrow argued that in complex and tightly coupled systems multiple unexpected interactions of failures are inevitable and will occasionally lead to an accident (Perrow 1984, p. 5). As such an accident is a system accident and as this behavior of the system is normal, Perrow called such accidents *normal accidents*. As Reiman et al. (2015) point out, adaptation is a vital feature of complex safety critical systems, but adaptation can also lead to system failure. Variance and fluctuations in the system may be the source of innovation and change, but also the source of accidents. Complexity has been called *the enemy of safety*. For Perrow, the two key characteristics were the *interactions within the system* which could be either linear or complex, and the *coupling* which could be either loose or tight (p. 5). Perrow (1984, pp. 85-86) characterized complexity in the following way:

- Proximity of parts or units that are not in a production sequence
- Many common mode connections between components (parts, units, or subsystems) not in a production sequence
- Unfamiliar or unintended feedback loops
- Many control parameters with potential interactions
- Indirect or inferential information sources
- Limited understanding of some processes

Perrow (1984, pp. 93-94) described tight coupling in the following way:

- Tightly coupled systems have more time-dependent processes, which cannot wait or stand by. In loosely coupled systems, delays are possible; processes can remain in a standby mode.
- The sequences in tightly coupled systems are more invariant.

- In tightly coupled systems, the overall design of the process allows only one way to reach the production goal.
- Tightly coupled systems have little slack. For example, quantities must be precise and resources cannot be substituted for one another.

These descriptions are probably colored by Perrow's very industrial context. He comes to the conclusion that systems which are interactively complex *and* tightly coupled have an intrinsic problem: because of their complexity, they should be decentralized, however, because of the tight coupling, they should be centralized. Perrow (p. 328) also discusses risk perception using so-called thin and thick descriptions of hazards. The former is a quantitative, precise, logical and value free hazard description. The latter takes a larger perspective and includes subjective dimensions and cultural values, typically being skeptic about human made systems and institutions and the ambiguous nature of experience. Interestingly, the thick description can be seen to resonate with the modern risk perspective and with the unpredictable nature of complex systems. Antonsen (2009) also quotes Perrow in the latter's response to the question why he did not discuss *culture* in Normal Accidents:

> "… Of course there are cultures in companies, but on issues of risk and safety I think the issue is really power. "

This quote takes us back to the forces and interactions within a complex system of human agents: an organization. A thin description does not capture a large part of the dynamics within a complex human organization.

The role of information is highlighted as very important within complex systems but also very problematic. Geographical dispersion and segregated knowledge about tasks means that real knowledge about what is happening elsewhere in the organization is often poor. Measures to increase the flow of information may easily result in an overflow of relatively useless information as opposed to facilitate real knowledge between skilled people (Reiman et al. 2015). In line with this, Snowden (2005a) warns that *context* is critical to knowledge and learning. He recommends implementing methods and techniques which accommodate informal communities and knowledge in narrative and concrete form, in contrast to relying solely on abstracted knowledge shared in formal communities. Snowden (2005b) also points out that significant aspects of what people know cannot be measured or made explicit: people always *know* more than they can *say,* and they can say more than they can *write* down.

Finally, it is important to introduce the concept of an *attractor*, and especially in the context of an organization, implying that the agents are humans and the attractors would need to be relevant for them. An attractor refers to properties toward which the system tends to evolve. The specific case of a strange attractor means that the exact values of the system in the attractor cannot be predicted. Examples include shared practices and values (Reiman et al. 2015). Anderson (1999) defines an attractor as a limited area in a system's state space that it never departs. In practice, as direct command is not a viable strategy for achieving wanted outcomes in complex systems, attractors often offer an indirect way to attract agents and ensembles towards the wanted behaviors.

4 Systems thinking

The objective of this section is not to make a comprehensive review of systems thinking with its different variations. Rather, the objective is to show the key concepts and methods that systems thinking has provided for understanding complex systems better. Other proposed strategies, concepts and methods for complex systems - beyond systems thinking - are introduced in the next section, the overall aim being to build an understanding on what could be the most promising approaches for dealing with complex systems: both for understanding what is going on in the system, and for being able to influence the system in a desired way – e.g. for risk treatment. The focus in this section is on the basic principles of the systems thinking movement. This movement can be associated with many different scientific and research strands, such as system dynamics, soft operations research and critical systems thinking. In

view of the current objective of focusing on the basic principles, detailed discussion of the different strands and their differences is out of the scope of this section.

The main strands of systems thinking emerged and became established in the period between the second world war and the early 1980s (Jackson 1994). According to Flood (2010), systems thinking emerged through a critique of reductionism. Checkland & Haynes (1994) argue that despite the differences in the different strands within the "systems movement", there is a reasonable degree of coherence within the movement as the different strands all make use of the concept of *system*, as an adaptive whole, an entity having emergent properties, a layered structure, and processes of communication and control that allow adaptation in a changing environment.

Systems thinking recognizes systems as such having properties worth studying and understanding. There is a conceptual framework built around the system concept and some typical methods and practices for studying system characteristics.

As stated in the beginning of this chapter, a system must consist of three kinds of things: *elements*, *interconnections* and a function or *purpose* (Meadows 2008, p. 11). To these can be added *events* (Senge 2006 p. 21) and system *behavior* (Meadows 2008, p. 88). As Senge points out, people are naturally drawn to events and try to find the causes for these events. This distracts people from seeing the longer-term patterns of change lying behind the events. These long-term patterns could also be called the system behavior. Like Meadows (2011, p. 89) writes: "system behavior reveals itself as a series of events over time". In terms of being able to influence a system, the biggest influence comes if one is able to modify the purpose. The next best opportunity would be at the level of the interconnections, while changing the elements is typically a much less promising strategy (Meadows 2011, pp. 16-17). The system behavior and the resulting events are visible but from the system point of view they are consequences of the purpose, interconnections and elements.

As stated above, the system purpose is not necessarily the one which is officially stated. It can be deducted from the system behavior, as the behavior reflects the real forces within the system. In line with the description of complex systems in previous sections, the systems are often nested one inside another, often in multiple levels. This also means there can be "purposes within purposes" and there may be conflicts between the purposes at different levels of the system (Meadows 2011, pp. 15-16).

In addition to being drawn by events, there are various system characteristics which make it difficult for people to understand how systems really work. Non-linearity (which has already been discussed above in the context of complexity) is an important factor here, and clearly recognized within systems thinking (see e.g. Meadows 2011, p. 92). People also tend to see systems more static than they are, when in fact most of the time systems are about *flows,* not about static situations. In systems thinking, a framework of concepts helps to work with these types of phenomena: for example, concepts of *stocks* and *flows* are used (see e.g. Meadows 2011, pp. 18-24). Feedback loops are naturally also very important, and a symbolism has been developed within systems thinking to draw systems diagrams, often consisting of a complicated structure of different (reinforcing and dampening) feedback loops. Several typical building blocks of complex systems have also been recognized and named as *system archetypes*. For an extensive description of the archetypes, see e.g. the appendix 2 of Senge (2006) and for good examples of system diagrams see Cooke & Rohleder (2006, p. 227) on the incident learning system, Marais et al. (2006, p. 573) on complacency related to safety and Leveson et al. (2006, p. 104) on the dynamics behind the loss of the space shuttle Columbia.

Starting from the idea that systems thinking is an alternative, opposing approach to reductionism, it is quite natural that in addition to *analysis*, systems thinking promotes *synthesis*. In analysis, the system under focus is first taken apart and the behavior of each part is studied separately. Finally, the understanding at the level of the parts is aggregated in an effort to explain the behavior of the whole. In synthesis, the system under study is first recognized as a part of one or more larger systems. There is then an effort to try to understand the function of the larger system(s). The understanding of the

larger system is then used to identify the role or function of the original system under focus (Ackoff 1999, pp. 11-12). For example, a safety department in an airline could be studied in terms of its capability to ensure safe operations. *Analysis* would focus on the different activities of the safety department and try to understand their contribution to safety. An approach through *synthesis* would see the safety department as a part of the airline and recognize the purposes and key drivers of the airline and its top management. This might reveal that in reality the weight given to the safety work, e.g. reflected in budgets and top management focus, is actually relatively low. Such a situation would obviously limit the capabilities of the safety department significantly.

Ackoff (1999, p. 13) introduces the concept of "mess". This somewhat provocative term is used to draw attention to the fact that problems should not be considered in isolation from each other. He argues that problems are almost never separable and that the typical situation consists of complex systems of strongly interacting problems. Ackoff calls such "systems of problems" *messes* and advises against the practice of reducing messes to lists of problems and then treating problems separately, as self-contained entities. From the systems thinking point of view, the mess needs to be considered as a whole.

It is difficult to define what would be a single unified position of systems thinking in terms of how predictable and controllable complex systems are. The different strands differ in their approaches. Some could be seen as having a lot of faith in excessive computerized modeling of complex systems (see Forrester 1971; and Meadows et al. 1982) or qualitative modelling (see Leveson 2011; and Goh et al. 2010). Meadows (2008, p. 169) argues that (complex) systems can't be controlled but they can be designed and redesigned. She states that complex systems are counterintuitive (p. 146) and that there are no quick or easy formulas for finding *leverage points* in complex and dynamic systems. It could be argued that assuming the existence of so-called leverage points where significant change can be achieved with minor effort, and entertaining the possibility that such leverage points could be located, assumes too much structure, stability and predictability in a complex adaptive system. Systems thinking could be abused as another attempt to try and *take control* of a complex system, even if this is not the original idea of systems thinking.

Snowden (2005b) criticizes systems thinking, and in particular systems dynamics, for maintaining the notion of centralized design and planning: "it assumes that focus and alignment is a top-down objective-based process". Snowden recognizes that system thinking embraces non-linearity and the inherent greater ambiguity related to human systems but argues that the "basic assumptions of order pertain".

However, as stated in the beginning of this section, the systems thinking movement consists of many different strands and while criticism on different aspects of these disciplines may be well justified, the focus here is on how systems thinking has enhanced the understanding of complexity and complex systems. The position adopted here is that systems thinking can be very helpful to counterbalance reductionism and to help understand complex systems better. Its various principles and methods (like system diagrams and archetypes) help illustrate key aspects of systems, such as feedback loops, nonlinearity and the importance of the system's purpose. It must be understood that all models and system descriptions are always incomplete and incorrect static descriptions of a system which in reality is typically even more complex and in constant evolution. It is also fair to say, that some aspects of complex systems can be illustrated with modeling, but this is far away from saying that such models can accurately predict the future.

5 Managing risks in Complex Adaptive Systems

What are then the implications of the scientific knowledge on complex adaptive systems on risk management and in particular risk treatment? The old Cartesian approach is clearly not the best way to deal with complex systems. Sagan (2004) gives a tangible example of this. Following a classic Cartesian

engineering approach, adding *redundancy* to the system increases its reliability and safety. Sagan shows how such redundancy can also have negative effects in a real-world complex adaptive social system:

- The system becomes more complex and can produce hidden common-mode errors
- Redundancy can lead to social shirking among humans in organizations. Humans are aware of one another and adding redundancy can lead others to be less observant or responsible.
- Redundancy tempts the management to increase production pressures, making the system perform at higher tempos or less safe conditions.

Similarly, Dekker et al. (2001) point out that barriers, professional specialization, policies, procedures, protocols, redundant mechanisms and structures all add to system complexity and entail an explosion of new relationships. System accidents result from relationships between components instead of a failure of a single component, and in the face of extensive complexity it becomes very difficult to understand and anticipate all the possible interactions.

Consequently, there is a clash between the old and new thinking of how to deal with emergent properties such as safety when dealing with complex sociotechnical systems. This paradigm change is similar to the old versus new risk perspectives discussed in Chapter 2, but perhaps even more profound. The objective of this section is to study the different proposed approaches for dealing with complex adaptive systems, especially in the context of safety and risk. It is acknowledged that the previous sections of the current chapter have already contributed to this discussion. The discussion starts with the Cynefin framework which can be helpful in recognizing when a particular problem/system is in the complex domain. After that, three specific challenges are discussed: *understanding* what is happening within a complex system ("input"), *learning* about the system, and trying to achieve *positive change* within the system ("output").

5.1 The Cynefin framework

The so-called *Cynefin* framework has been proposed to give clarity on what kind of system one is faced with, and consequently what types of approaches could be the most adapted (Kurtz & Snowden 2003, Snowden & Boone 2007, Sturmberg & Martin 2008, Hasan & Kazlauskas 2009, Gorzen-Mitka & Okreglicka 2014, French 2012, 2015). The phraseology used within the Cynefin framework has evolved over time and may still evolve. The framework is presented in Figure 13.

According to Kurtz & Snowden (2003), Cynefin is a *sense making* framework meant to help people in decision-making. There are five areas within the Cynefin. On the right side, there are the *ordered* systems: *obvious* and *complicated*. On the left side, there are the *unordered* systems: *complex* and *chaotic*. Finally, there is a small area in the middle called "*disorder*" which symbolizes the case where it is impossible to know where the system in focus belongs. The simplest case is the obvious domain. The cause-and-effect relationships are generally linear and as the name suggests, obvious to everybody. Repeatability allows predictive models to be used. Therefore, in this domain predefined best practices can be used, e.g. operational procedures. The focus can be on efficiency. The decision model is to categorize the incoming data and respond in line with predefined best practice. Other names for this domain (used previously) are "simple" and "known".

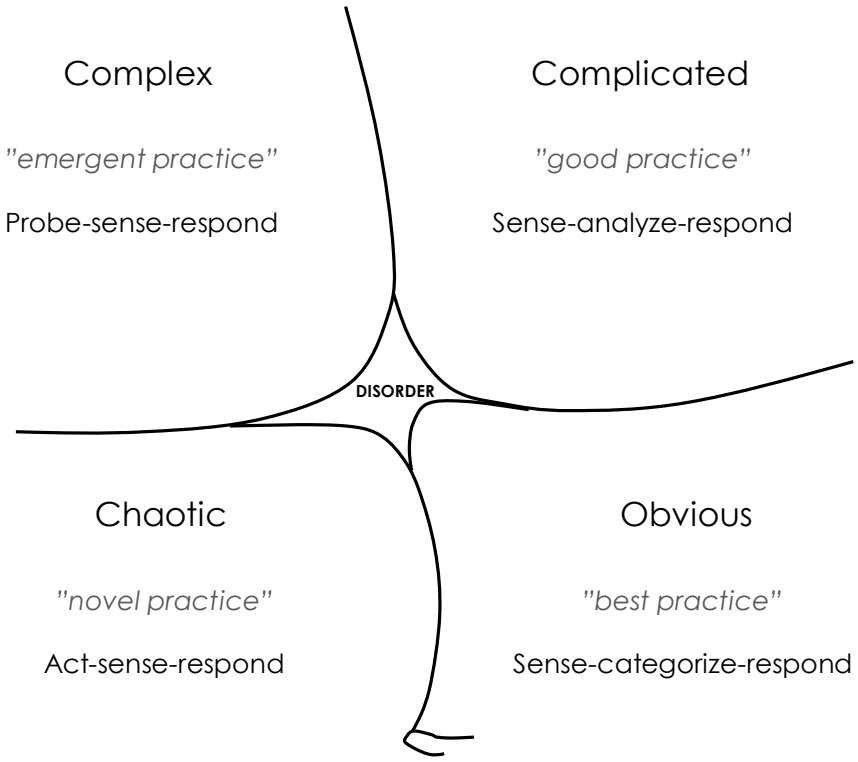


Figure 13. The Cynefin framework. Based on Kurtz & Snowden (2003) (with updated terminology by Snowden from www.cognitive-edge.com).

The meaning of complicated and complex within the Cynefin framework are well in line with what has been presented in the previous sections. In the complicated domain, there are still cause and effect relationships, but figuring them out is now more challenging than in the obvious domain and usually requires expertise and resources. In theory, everything in this domain could be moved to the obvious domain if a sufficient effort was made. However, in practice the usual solution is to rely on expert opinion. Instead of one best practice, there are typically many good practices and analysis is the key for finding the most appropriate solutions. This domain could also be called the domain of "knowable".

The complex domain of the Cynefin framework is about everything that has been reviewed concerning complex systems in the previous sections: high number of relationships defying categorization or analytic techniques, emergent behaviors and phenomena, unpredictable futures, knowledge within the system being scattered all around the system, etc. Kurtz and Snowden (2003) stress that the methods, tools and techniques of the obvious and complicated domains do not work here. According to them, in this domain, system behavior may seem predictable when observed afterwards: this phenomenon is called retrospective coherence. Once a pattern has stabilized, it can indeed appear logical, but many other patterns could have stabilized too, and seem equally logical. There is no guarantee that observed patterns continue to repeat, and the underlying sources of the patterns are not open to inspection. Relying on expertise (which works in the complicated domain) easily leads to solutions which address the historical patterns but do not help recognize and cope with new and different patterns. They recommend the use of probes to experiment with plausible solutions. The idea is to make patterns (or potential patterns) more visible before decision making. Desirable patterns observed in successful probes are strengthened and undesirable patterns/probes and dampened down. Kurtz and Snowden also refer to techniques using narratives in order to capture the knowledge within the system. They argue that understanding this space requires to gain multiple perspectives on the nature of the system and to feed such insights to the decision makers.

In the domain of chaos, there are no visible relationships between cause and effect. In this domain, it is necessary to act decisively in order to reduce the turbulence. The domain of disorder in the center caters for situations where it is difficult to know which domain would be applicable, or when different people disagree on the domain.

A system does not necessarily stay within one domain of the framework. For example, the bottom of the framework symbolizes a potential drop from the obvious domain into the chaotic domain: such a catastrophic failure could happen if the complexity of the situation was underestimated and the signs of an emerging change were missed (Snowden & Boone 2007). There are also desirable movements from one area to another, such as letting the system go into the chaotic state for a limited time in order to remove some constraints and enable innovation and renewal. Also, an activity in the complicated domain may be mastered so well, that it can be standardized and moved to the obvious domain. Kurtz and Snowden (2003) point out that a move across boundaries requires a shift to a different model of understanding and interpretation as well as a different leadership style. For more details on the Cynefin dynamics, see Snowden (2005c) and for applied examples, see Sturmberg & Martin (2008) and Beurden et al (2013).

5.2 Understanding what is going on within the complex system

As discussed in previous sections, one of the key features of a complex system is that information and knowledge are distributed everywhere within the system. It therefore becomes a challenge for anyone to understand what is going on within the system. As the system is open, the environment is in constant interaction with the system and is therefore another "part" of the system to monitor.

Inayatullah (2002) points out that the target has to be to learn about "things we don't know about" but also about things "we don't even know we don't know about". He argues that multiple approaches are needed and proposes three necessary steps:

- Ensure that there is an environmental futures-scanning process
- Ensure that the process uses multiple methods
- Ensure that data/insights come from arenas outside of official power (also outside official formulations of what is normal or acceptable).

As pointed out in Chapter 3.3, the solution should not focus on abstracted information (or "data") only, because context is so important to real knowledge, and attempts to increase information flow may easily turn into an overflow of relatively poor information (Reiman et al. 2015, Snowden 2005a). Snowden (2005a) also promotes relying more on informal communities, and knowledge in narrative and concrete form. There is agreement that multiple *different perspectives* are necessary for trying to understand the events and evolutions (e.g. Mikulecky 2001, Dekker et al. 2011) within complex systems, as illustrated in Chapter 3.3. These points suggest a lot of emphasis should be put on the people and on knowledge sharing between people, as opposed to, and in addition to, "data transfer" and computerized methods.

5.3 Learning about system characteristics and behavior

As Sterman (1994) states as one of his conclusions on learning "in and about" complex systems, "complex dynamic systems present multiple barriers to learning". He argues (p. 310) that to learn effectively in the world of dynamic complexity and imperfect information, people must develop special insight skills that help people learn when feedback is ambiguous. He suggests that effective learning can be based on *continuous experimentation in both virtual worlds and real worlds* and feedback from both. The feedback should inform development of mental models, the formal models, and the *design of experiments*. The virtual worlds can be physical models, role-plays or computer simulations. They can be used as low-cost laboratories for learning, where time and space can be compressed or dilated, actions can be repeated under the same or different conditions and one can stop the action to reflect when necessary. The system can also be pushed into extreme conditions which can reveal more about its structure and dynamics that incremental adjustments. One can experiment with dangerous situations and even catastrophes. Sterman (1994, p. 318) has built a model of the idealized learning loops related to

learning about complex systems through both real-world and virtual world. He argues that simulation is the only practical way to test our mental models which are usually characterized by too narrow temporal and spatial boundaries – they are dynamically deficient, omitting feedbacks, time delays, accumulations and non-linearities (see Chapter 4.3 for discussion on the various biases affecting humans in this context). However, Sterman also warns about many pitfalls in this kind of learning, including not taking time to reflect on the outcomes, defensive routines and groupthink, and the challenge of learning when facing long time delays, as well as causes and effects which are distant in time and space.

Sornette & Ouillon (2012) support the simulation approach for learning. They argue that progress comes only by living through scenarios and experiencing them.

Based on the nature of complex systems, as discussed in previous sections, one needs to keep in mind the limitations of simulations. No model will ever replicate the real system in a perfect way and it will be very difficult to replicate the constantly evolving and co-evolving nature of complex systems. Many of the human-related aspects (e.g. the informal channels of transferring knowledge) of these systems will also be extremely challenging for modeling.

As seen in the previous section, designing experiments and carrying out continuous experimentation are supported by the Cynefin framework, where the main approach in the complex domain is to use probes and then adapt further responses based on the experiences from the probes.

Carrying out experiments in a complex system should not be confused with *controlled* experiments. As Sterman (1994) points out, due to the dynamic complexity of complex systems, in the best case controlled experiments would be prohibitively costly or unethical, but more often *it is simply impossible to conduct controlled experiments in complex systems*. As already reflected in the previous sections, complex systems are in disequilibrium and evolve continuously, many actions yield irreversible consequences, the past cannot be compared well to current circumstance, many variables change simultaneously and there are multiple interacting feedbacks so it is difficult to imagine how some aspects of the system could be held constant to isolate the effect of the variable of interest, or how to interpret the changes in the system behavior in the existence of all these inherent system characteristics. As Cilliers (1998, p. 23), Le Coze (2005) and Delorme & Lassarre (2014) point out, this becomes a problem if one wants to follow a very classic scientific approach where general laws are formulated and then empirically tested in the real world through controlled experiments and observation, following "the correct method". When dealing with complex phenomena, no single method can yield the whole truth and due to the intractability, the inquirer's role is no longer backed up by an objective technique.

One perspective to learning is to relate to the DIKW-hierarchy. As Rowley (2007) presents, the most common version of this hierarchy consists of four layers: data, information, knowledge, wisdom. The rationale behind this hierarchy is that data can be used to create information, information can be used to create knowledge and so on. The highest level in the hierarchy is wisdom, and as every level in the hierarchy contains all the levels below, wisdom contains the three other elements in hierarchy. Ackoff (1989) likes adding a fifth layer called understanding. These five levels can be defined in the following way (Rowley 2007, Ackoff 1989, 1999):

- Data are symbols that represent properties of objects, events and their environment.
- Information is contained in descriptions and answers questions beginning with *who*, *what*, *when*, *where*, *how many*?
- Knowledge is know-how and enables transformation of information into instructions. Knowledge answers the question *how to*? Knowledge can be obtained from other people, by instruction, or by extracting it from experience.
- Understanding, according to Ackoff, is contained in explanations, answers to *why* questions. It is required for determining the relevance of data and information and to relate a situation causally to the objectives. Understanding facilitates and accelerates the acquisition of knowledge.

- Wisdom is the ability to perceive and evaluate the long-run consequences of behavior. It is the ability to increase effectiveness. It requires judgment which again implies ethical and aesthetic values which can be considered unique and personal to the actor. According to Ackoff (1989), development cannot occur without wisdom, because development is associated with increase in *value* (which is not necessarily the case for growth).

Rowley (2007) examines the hierarchy from different perspectives and presents among other things that:

- At the lower levels of the hierarchy, solutions can be *algorithmic* and *programmable,* but this becomes less and less possible when one moves towards wisdom.
- Meaning, applicability, transferability, value, human input and structure all increase as one moves higher in the hierarchy.

Ackoff (1989) is in line with these two points when stating that we will never be able to generate wisdom by computerized information-, knowledge-, and understanding-generating systems.

5.4 Achieving positive change in a complex system

Based on the previous sections, it is clear that one should not dream of *controlling* a complex system in the classic top-down sense. Such control does not exist, and even the necessary knowledge about the system is spread all over the system without anyone having the full picture.

As an example, in their (financial domain) paper "living in a world of low levels of predictability", Makridakis & Taleb (2009) recommend avoiding the illusion of control, recognizing that accurate predictions and controlling future outcomes are not possible, and acknowledging uncertainty - including the possibility of black swans.

Furthermore, experience on human control of complex systems is not encouraging. Sterman (1994) characterizes human performance as "far from optimal" and Sornette & Ouillon (2012) state that humans experience failure more often the success when intervening in complex systems. Some of the contributing factors in terms of complex system characteristics have been discussed above and the specific problems related to human biases and human performance limitations are discussed in Chapter 4.3. The objective of the rest of the current section is to review advice in terms of strategies for dealing with complex adaptive systems. These strategies can be taken into account both at the level of the organization carrying out the risk assessment (e.g. national transport safety agency) and at the level of the various organizations which are part of the complex operational system under study (e.g. the transport system in a country).

From this theoretical point of view, one can start with the hierarchical elements of complex systems as reviewed in Chapter 3.4, listed in the hierarchical order:

- Purpose
- Interconnections
- Elements
- Behaviors
- Events

In terms of influencing the system, if one succeeds in modifying something related to the top of the list, the influence will be much greater than with the items at the bottom of the list. A practical example could be a manager in an organization who does not put a lot of weight on safety: replacing the person might not have a lot of effect if the new person is faced with the same organizational pressures and incentives as the previous one. However, if the real purpose changes, implemented through personal incentives etc., the behavior of the manager is bound to follow the new purpose.

Many of the points raised by Marion & Uhl-Bien (2001) on complex leadership (see Chapter 3.3), can also be referred to when designing purposeful interactions with a complex system, as these points are based on fundamental characteristics of complex systems. For example, considering the nested nature

of complex systems, the following elements could be brought from the intra-organization level of departments and workgroups to the system-level of interacting organizations and interest groups (sometimes with minor adjustments):

- Creating the conditions that enable productive, but largely unspecified, future states; as well as emergence of *distributed intelligence*.
- Influencing networks and creating atmospheres which catalyze the formation of aggregates and meta-aggregates and emergence of innovations and their dissemination (now understood as between organizations instead of between individuals or work teams)
- Cultivating largely undirected interactions among organizations and other groups of people (e.g. people sharing a common hobby) to enable common understanding and the creation of favorable uncontrolled futures and productive surprises.
- Understanding organizational behavior in terms of *global interactions* rather than narrow focus on controlling *local events*.
- Enabling effective *networks* rather than trying to *motivate* enhanced effort or keep peace.
- Using interaction to enable individuals and workgroups (and organizations & interest groups) to work through *conflicting constraints.*
- Learning to manipulate the *situations of complexity* rather than the results.

Snowden (2015b) warns against trying to apply approaches appropriate for ordered systems in the context of complex systems and argues that the appropriate strategies are manipulation of boundaries, attractors and identity. He refers to:

- Making decisions based on patterns arising from personal experience and collective knowledge, expressed in narrative form
- Managing starting conditions and monitoring for the emergence of patterns
- Allowing for inefficiencies for the benefit of adaptability and effectiveness
- Instead of measuring outcomes compared to goals, trying to measure the stability of barriers, the attractiveness of attractors, and the stability of identities.
- Being able to move towards successful states, without being able to define in advance (except in most general terms) what the success would look like.

The point Snowden (2005a) makes about informal communities (which was referred to in Chapter 3.5.2) may also be relevant for having an influence on the system: the point is not to focus only on the formal structures and processes.

Sterman (1994) warns against imagining that people in an organization behave in an ideal way. He mentions local incentives, asymmetric information and private agendas as factors which can lead to game playing by agents throughout a system – this is linked with the topic of power discussed in Chapter 3.3. He also refers to defensive routines which may lead to mental models of team members remaining hidden, ill formed, and ambiguous. This can be associated with groupthink where members of a group mutually reinforce their current beliefs, suppress dissent, and isolate themselves from others with different views or disconfirming evidence.

As discussed in the previous section, an approach which seems to naturally fit in the context of a complex adaptive system is to carry out experiments or "probes" in the Cynefin terminology. As the idea is not to use a single method or to test a single course of action, this suggests that experiments should rather be used to explore the system and its behaviors through a number of different experiments. The idea is to adapt the course of action based on the results of the experiments and carry on further experiments.

6 Synthesis and conclusions

The transport system is a complex adaptive system. State agencies, commercial operators, interest groups, urban transport providers, private people using public transport or their own vehicles, etc. are elements in the system. They have a huge number of interactions in terms of communications, cooperation, business transactions, conflicts, people moving from one organization to another, etc. The

presentation of the Finnish maritime cluster in Figure 14 is a good illustration of this: this is only a part of the transport system in the maritime domain, but it already constitutes a complex network of interacting organizations, each with their specific interests and constraints.

The system is very open: for example, conflicts in distant parts of the world have an impact on the oil price, which is then reflected in the price of tickets for transport; this again can change the competitive set up locally. Urban interest groups may put pressure on airports to reduce noise, which may contribute to less safe takeoff and landing procedures at airports (see Eurocontrol 2014, pp. 34-35 for a great example). The system has a huge number of feedback loops, for example regulatory controls limiting reckless operations, and the dynamics of offer and demand within the transport business. The system is constantly evolving, for example with increasing pressure for environmentally sustainable transport options and with some emergent new phenomena like low-cost airlines, BlaBla cars and Uber. The last two examples remind of the fact that the development of internet-based solutions has had a drastic impact on the transport system – another feature of the open system. This also offers an example of co-evolution: making it possible for the customers to manage their bookings online is bound to contribute to evolution where more and more people use such a service but may also feed back as growing demands for more and more options to be available online and for more and more transport providers to have a strong online presence. The transport system is also a nested one (team-department-airline-airline alliance/community of operators in a country- international community of airlines- aviation community) and a system where the same element can be part of many overlapping subsystems and change from one role to another: the commercial pilot drives his private car after work and becomes a pedestrian the moment he/she leaves the car.

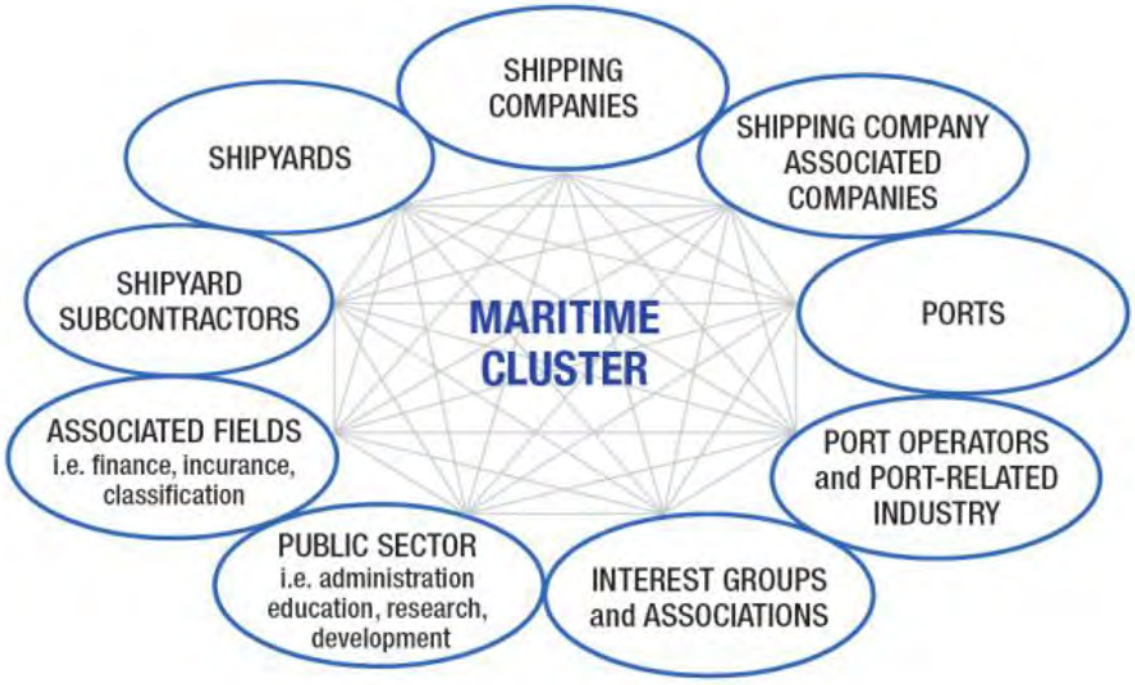


Figure 14. The Finnish maritime cluster (Aalto university 2012).

Because most parts of the transport system are organizations, all the typical features of human organizations discussed above also apply, e.g. informal structures and power issues.

The first conclusion is thus that the transport system should be treated like a complex adaptive system, in all interactions with the system, e.g. when trying to learn about system behavior and when trying to influence the system positively. However, accepting the unpredictable, counterintuitive and

uncontrollable nature of such systems is not the dominating current practice. Methods adapted for dealing with complex systems - such as experimenting, using information from non-experts, relying on narratives or informal social structures - easily seem counterintuitive and even unprofessional. While shifting to the new paradigms and methods and trying to understand the dynamics behind the system behavior, one should not create the illusion that these methods (e.g. system diagrams) will provide full understanding or control of the system. While patterns may be common in the behavior of some complex systems, the inherent unpredictability will remain.

In terms of the practical methods that could be used, a good starting point is the Cynefin framework, as it helps to map which systems or which aspects of the system may be on the complex domain. System diagrams can also be helpful in highlighting some of the interactions within a system, even if these diagrams will typically be only harsh drafts of the reality. Even so, the diagrams may help in resisting the simplistic Cartesian view. While not ruling out the possibility of using computerized simulations, it is acknowledged that the transport system is so complex that building a useful computerized model could turn out to be nearly impossible or at least an overwhelming effort. The scope of the model would probably need to be at the level of the whole society (of a nation).

As information and knowledge is distributed all over the system, it is important to involve a diverse group of people - of experts and non-experts - to get different perspectives to what is going on. As things are interconnected, problems are also interconnected into "messes" and should not be treated in a fragmented way. Similarly, safety cannot be treated in isolation from other aspects of the system as it typically interacts with other priorities. In terms of interventions, direct action can be considered risky as the system responses are very difficult to predict and there are typically unintended consequences. Therefore, small parallel experiments which can be scaled up or dampened down are safer intervention strategies.

The second - somewhat surprising - conclusion is that despite the many obvious shortcomings related to human cognition and human behavior in groups, humans possess qualities which are invaluable in the context of complex systems and which cannot be replaced by computerized systems or abstracted data. To try to maximize the understanding of a complex system several diverse perspectives from different disciplines are necessary; information must be kept within its context which - among other things - favors the narrative form and passing knowledge directly from a person to other(s); to learn at the level of wisdom requires understanding objectives, values and long-term consequences holistically, and this involves subjective value judgments; in organizations, emergent properties like safety rely on people; and knowledge transfer relies partly on informal undocumented human contacts. Computers are not able to take different perspectives or to apply different scientific disciplines for interpreting events, neither are they good at understanding multiple important nuances in narratives or making ethical judgments, etc. This does not mean that technical artefacts, computerized applications and databases are useless: the goal could be to build support systems which enhance the use of the positive qualities of humans and if possible decrease the tendency to fall under the influence of the typical human biases. Consequently, emphasis is put more on people than in most state-of-the-art practices.

The third conclusion is that due to the fundamental unpredictability of complex adaptive systems, in the context of risks, many uncertainties must be considered not only *unknown* but also *unknowable*. This kind of uncertainty is not necessarily related to lack of knowledge: in a complex system it is already very difficult to imagine *what* could happen, not to mention related probabilities.

The fourth conclusion is related to the impossibility of controlled experiments. This means that causality between two things may become practically impossible to prove. Consequently, the scientific method in its original form relying on controlled experiments and observations becomes severely handicapped in complex systems. This point needs to be taken into account in Part III, where the aim is to evaluate the developed risk management framework and to compare it with the set objectives.

Besides the conclusions which feed the requirements for the NRMF, the topics reviewed in this chapter form an important body of knowledge for people involved in risk treatment in a CAS.

Finally, another important observation is that safety - which can be considered the flip side of risk - is an emergent property of the system - in this case, the transport system. It cannot be created by someone commanding it from the top, nor through a prescribed program: specific histories make organizations different with different responses to similar circumstances and lead to distinct emerging organizational cultures. Likewise, most accidents in complex systems can be considered emergent. Safety and the evolution of safety management strategies are the topics of the next chapter.

Chapter 4 – Contribution of Safety Science to Risk Management

This chapter covers the evolution in the understanding of safety and in safety management strategies. As safety has become entangled with *human factors* - especially in safety critical activities with human operators, like aviation - the end of the chapter contains a brief review of some key human factors aspects, which also gives the opportunity to summarize some of the relevant human biases and limitations. The safety topics covered in this chapter are directly linked with risk management. Increasing safety could be understood as a synonym for lowering risks. Therefore, it is worthwhile reviewing what safety science can propose in terms of paradigms, strategies and methods.

In 2006 Erik Hollnagel, David D. Woods and Nancy Leveson edited the book "Resilience engineering - concepts and precepts". This book could probably be considered the landmark for a new approach to safety called *resilience engineering*. Resilience engineering was born to offer something that could replace the classic approach of safety as the absence of negatives. Safety had to be the *presence of something*. To mark the difference between the two approaches the terms *safety-I* and *safety-II* were created, the former referring to the classic approach and the latter referring to the resilience engineering approach. In 2014, Erik Hollnagel published the book "Safety-I and safety-II - the past and future of safety management" focusing on the differences between the two approaches.

Safety data – e.g. safety events - are a key input to the risk management process and the available data and the way they are processed (e.g. categorized) depends on the existing safety management practices. Available analyses and safety indicators are inputs for risk identification and can also play a role during risk analysis. In this way, the classic world of *Safety-I* contributes to the risk management process. Even the Reason model from 1990 has its importance, as methods like ARMS (see Chapter 6) carry out risk assessment based on the barrier-thinking of the Reason model. The more recent *Safety-II* paradigm with its focus on resilience is an important concept in risk treatment because it is more promising to address (especially the low probability) risks by increasing system resilience than by trying to address specific scenarios one by one. The question to ask is, what does safety science recommend for increasing the resilience of crews and organizations? The Safety-II review is concluded by discussing some of the most important current challenges and paradoxes in safety management. The main interest for this topic is in creating a strong theoretical basis for effective risk treatment.

As "fixing the human error problem" was seen as a primary risk treatment strategy for a long time, it is important to set the record straight and propose an up-to-date understanding about human error and (operational) human factors in general. Finally, the human biases, heuristics and other such limitations have a direct impact on risk perception, decision making and the capability of a group to work effectively on risk analysis and evaluation - thus affecting several stages of the risk management process.

1 Safety I: safety as lack of accidents

The classic way to think about safety is to see it is as a lack of negative outcomes such as accidents. Aven (2014b) refers to several definitions for safety, in line with this thinking:

- Safety is the condition of being safe; freedom from danger, risk or injury
- Safety is the condition of being safe from undergoing or causing hurt, injury or loss
- Safety is the absence of accidents, where an accident is defined as an event involving an unplanned and unacceptable loss

He also points out that even if an organization has not experienced accidents, we should not refer to high or low safety but rather speak about the probability or uncertainty about its safety, as the future events

and their consequences are still unknown and could materialize. There is thus an inherent difficulty with this definition unless if a purely historical and statistical approach is adopted, but this would assume the future to be an unchanged continuation of the past.

Hollnagel (2014, p.1) offers a more detailed generic definition of safety:" safety is the system property or quality that is necessary and sufficient to ensure that the number of events that could be harmful to workers, the public, or the environment is acceptably low". Dekker (2003) calls this view "safety as the absence of negatives".

From the practical point of view, definition of safety as lack of accidents is very understandable as the accidents are the visible outcomes and the outcomes that everybody certainly wants to prevent. Not surprisingly, this has also influenced the approach to how safety has been addressed classically, focusing on accident prevention. Such a view is coupled with the classical understanding of causality, reductionism and other aspects of the Cartesian worldview, as discussed earlier in Dekker et al. (2011). The reductionist drive to split dynamic events into smaller and smaller factors is easy to see in the taxonomies used to categorize incidents and accidents, still today: for example, in civil aviation, just one part of the ECCAIRS taxonomy, the so-called descriptive factors, contains a list of 102 pages of factors, including 8 pages to cover different types of failures related to flight crew's judgment and operation of the various aircraft systems, such as (ECCAIRS 2010):

- Flight crew's operation of autoflight system
- Flight crew's operation of brakes
- Flight crew's operation of carburetor heat
- Flight crew's operation of door system
- Flight crew's operation of electrical system
- Flight crew's operation of emergency brakes
- Flight crew's operation of miscellaneous equipment
- Flight crew's operation of navigation lights
- Flight crew's control of the aircraft
- Flight crew's control of the aircraft's airspeed
- Flight crew's control of the aircraft's altitude

Lee and De Landre (2000, p. 10) provide an example of the inherent problem of the reductionist approach in safety management tools, related to the Australian national aviation safety reporting tool (OASIS) at the time:

> "A review of the usage of the existing OASIS events and descriptive factors was completed using data from 1993 to 1995. This revealed that of the 1400 descriptive factors available in OASIS, only 50% were used on average in a year, and that 29% were not used at all. It was further found that if 75% of the least used descriptive factors were removed, it would impact on only 0.5% of occurrences."

The focus in accidents, the dominating cartesian worldview and the unreliable nature of technology in the early days of aviation, for example, made a natural fit. Indeed, As Nisula (2014a) argues, focusing on cataloging and counting failures could work well in a system where the main safety driver is unreliable technology. However, in the last 100 years the reliability of technology has improved dramatically, and together with other factors has contributed to some industries becoming what Amalberti (2001) calls ultra-safe systems. Transport has been particularly benefiting from better technology, making commercial civil aviation and train transport in many countries ultra-safe, having major accidents only every few *million* operations (Amalberti 2001). As Hollnagel (2004, pp. 45-47) argues, gradually accidents were explained less and less as technical failures and seen more and more caused by human performance or organizational factors.

1.1 Incident reporting and chasing the human error

Amalberti (2001) argues that in most industries safety is governed by three simple and self-explicit principles:

1. Conceptual designs generate systems with a high theoretical safety performance but these systems are subject to technical and human failings. Such failings are noise in the system and should be totally eliminated or at least minimized.
2. As technology is increasingly safe, focus is on further reducing *human errors*.
3. *Safety reporting* is fundamental for improving safety. The reporting highlights the undesirable noise in three areas: technology, the human operator and the resulting situations including situational, organizational and systemic failings.

As the safety level has improved there are less accidents to use as source material for further improvement. At the same time, in line with the third point of Amalberti, safety reporting has created a flow of operational safety incidents. Focus then shifted on incidents with an underlying - perhaps implicit - theory that accidents could be prevented through working on incidents.

Safety reporting has become one of the pillars of safety management especially in safety critical domains, such as commercial civil aviation and nuclear power production. The regulators have also been active in this field, both by mandating some reporting and through supporting the establishment of reporting programs (Johnson 2003, p. 21). Over the years, a lot of know-how has been gained on how reporting systems can be run - as illustrated by the 1000-page handbook of accident and incident reporting by Johnson (2003).

Johnson (2003, pp. 22-23) identifies seven arguments justifying the development and maintenance of incident reporting systems:

1. Incident reports help find out what prevented the accident, e.g. by identifying the effective barriers.
2. The higher frequency of incidents (compared to accidents) permits quantitative analysis.
3. Incident reports remind of hazards thus increasing the likelihood that recurrent failures are noticed and acted upon.
4. The reporting system provides a way to keep the staff involved in safety improvement if the material is fed back to the staff.
5. Data and lessons can be shared and can be used for comparisons between organizations or industries.
6. Incident reporting schemes are cheaper than the costs of an accident.
7. There may be a legal/regulatory requirement to carry out incident reporting.

Furthermore, based on the conclusions on dealing with complex systems in the previous chapter, it could be pointed out that the typical narrative format of incident reports can be considered particularly suitable for transferring knowledge within its context in a complex system.

A central question related to working with incidents is how much does it contribute to preventing major accidents. Johnson (2003, p. 24) refers to several studies which seem to support the transferability of conclusions from incident reporting to accident prevention. There are also opposing views. For example, Hale (2001) points out that there seems to be a profound belief in the safety world that major accidents and incidents (Hale talks about “minor accidents”) have the same causes and should be prevented using the same measures. Interestingly, he goes to the roots of the often-used “safety iceberg” - a triangle where the peak represents (the few) high-severity accidents and the low levels represent (a high number of) incidents of decreasing severity. He points out that this so-called Heinrich triangle was created in the context of occupational safety, for example counting various levels of injuries taking place when personnel crossed a railway track. According to Hale, there was never a claim in the original texts that the underlying causes for each degree of seriousness were the same. He comes to the conclusion that different event types all need their own pyramids and if such pyramids were combined together the

resulting structure would not really be a triangle anymore, as the lowest level ("deviation and recovery") would be far larger than the mean level ("near miss") which again would be too large to fit nicely to the top part. He also argues that not all bottom level deviations have the potential to lead to loss of control (and thereby to major accidents) and postulates that the more sophisticated the preventive system becomes, the less similar will be the causal sequences of minor and serious accidents.

The role of incidents is important in the risk management process, as they can be used to create the primary flow of operational safety data as an input to the process. Based on the above, one can assume that the relationship between accidents and incidents is not straightforward but that there is obviously a link. It is not sure that all incidents could have escalated into accidents, but incidents point to weaknesses in the system.

As the second point in the list of Amalberti (2001) indicated, one major phenomenon in the classic safety management has been the strong focus on *human error*, and on human factors in general. He writes:

> "This human error reduction concept was extensively explored for some 20 years. Research funds were first spent on studying human reliability in engineering sciences, the human component being considered as an additional element in the system, similar to other technical components."

Some of the most relevant results from this human factors research will be summarized in Chapter 4.3. Dekker (2003) sees the "error counting" approach as a part of the "safety as the absence of negatives" mindset and reasons that such approaches are appealing to the industry for the same reasons that any numerical performance measurements are: error counting can be used as a quantitative basis for managerial interventions - no matter how bad such a practice may be. One of the important milestones in understanding human error was the book titled "Human error" by professor James Reason (1990). In his book, Reason presented a review of research in human error and proposed a classification of human errors. Importantly, in the chapter 7 of the same book, James Reason took the discussion to another level by taking a systemic view to accident causation at the organizational level and presented his famous model of accident causation, known as the "Reason model" or the "Swiss cheese" model. The Reason model is still today probably the most important model in the domain of safety management in terms of its influence on safety thinking since its introduction. It is argued that even when this model has not been explicitly mentioned, it is often implicitly present in the form of an underlying theoretical framework.

1.2 The Reason model on accident causation

Since the creation of the model in cooperation between James Reason and John Wreathall (Larouzée & Guarnieri 2015) and it being published in 1990 in the book *Human Error*, the model has been refined several times (see e.g. Maurino et al. 1995, Reason 1997). Compared to models presented thus far it introduced two new important dimensions: the *organizational* perspective and the focus on so-called *latent failures* which could take place far away from the accident both in time and distance. The model can be presented graphically in several simple formats - due to its many aspects even Reason himself used various different presentations to illustrate the model. The model can be linked with sub-frameworks like Reason's taxonomy on human error.

The impact of the Reason model has been significant. Larouzée & Guarnieri (2015) refer to 25000 citations according to Google Scholar and list examples of sectors applying Reason's work: aviation, marine, healthcare, defense, nuclear, oil &gas, rail and road safety. The Safety Management Manual of the International Civil Aviation Organization from 2013 still presents Reason's model as the main framework for understanding accident causation. The human error classification from Reason is also presented (ICAO 2013). The Swiss cheese and latent failure concept also still appear on the 2014 edition of the Federal Aviation Administration's Safety Management Manual for air traffic organizations (FAA 2014). It is a fairly safe assumption that most of today's safety work is still at least partly based on the approach that James Reason laid out in 1990 both in terms of what is written down and the often-implicit

models that people have in their minds. A typical example is the common practice to think in terms of "barriers". Appendices 1 and 2 show that the model underpins risk assessment methods too. The latent failure term and the various visual presentations of the model helped to put more weight on the organizational aspects than was the case in most other methods, like fault trees and event trees presentations where an accident is essentially seen as a chain of events (see e.g. Sklet 2004, Rasmussen & Svedung 2000, pp. 19, 33, 36).

Perhaps the most common way to present Reason's model is to show how the accident trajectory goes through the holes in several layers of barriers (Figure 15). Other versions of the model can be found in Maurino et al. (1995, p. 24) and Reason (1990, p. 210). The general idea is that safe systems have several layers of defences in place to protect the operation against accidents. However, none of the defences are perfect and they may also deteriorate in time - thus the holes. These deficiencies evolve in time and often it is the activities and priorities of the organization itself which take their toll on the defences. Sometimes the simultaneous existence of several deficiencies in several layers of defences enable the accident to take place - the holes align and the accident "trajectory" passes through all the defences.

Only the last layers relate to acts by the humans who actively control the system, for example the pilots or drivers. Therefore, the model moves the focus from solely looking at the errors of the frontline operators to examining the whole organization and the way its different layers contribute to safety and sometimes to accidents. The other layers describe organizational factors which may contribute to the accident in different ways. Chronologically the first layer refers to decision-makers at the level of the organization and corporate management. Through these layers, Reason introduces the concepts of active and latent failures. The following paragraphs discuss these elements in some detail (Reason 1990, 1997, 2000, Maurino et al. 1995, Larouzée & Guarnieri 2015).

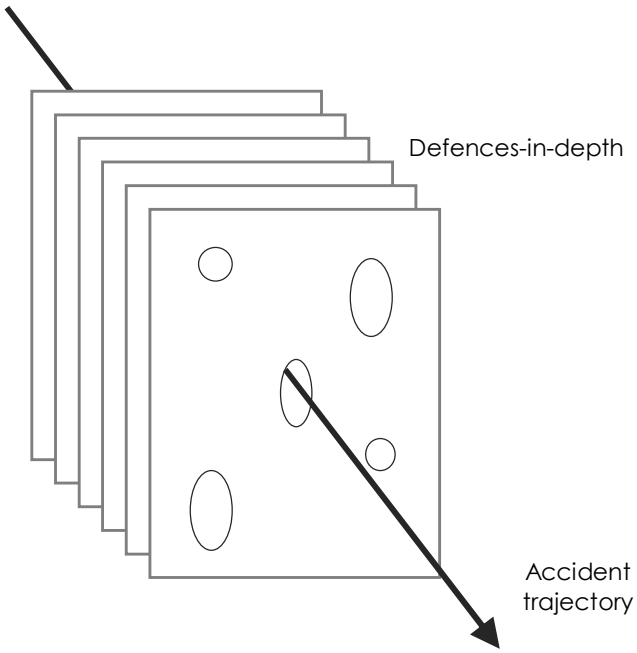


Figure 15. Presentation of the defences-in-depth or the "slices of Swiss cheese". Adapted from Maurino et al. (1995, p. 25).

Active failures are tangible acts committed close to the accident in time and place. These are the visible errors or omissions that are too easily offered as the "cause" of the event - for example "pilot error". *Latent failures* or *latent conditions* (later evolution of the term) are failures without immediate visible consequences. They are often committed far in time and place from the resulting event by people not

necessarily in direct connection with the aircraft, ship, train or car. Latent conditions decrease the safety margins in the system, making the system more vulnerable to accidents. The *event* in Reason's model is an unwanted operational occurrence, e.g. an accident. According to Reason accidents occur typically through a combination of active and latent failures. Sometimes the event does not simply occur as a result of the active and latent causal failures but also requires one or more contextual conditions to be fulfilled. Such conditions can be called *local triggers*. Such conditions are not causal factors to the accident but rather triggers which are necessary (but not sufficient) conditions for the specific accident to take place, e.g. the cold night affecting the O-rings before the launch of the space shuttle Challenger (Reason 1990, p. 254).

Latent failures are closely linked with the so-called *defences, barriers and safeguards*. Maurino et al. (1995, p.10) refer to *defensive measures aimed at removing, mitigating or protecting against operational hazards*. They present two ways to classify defences: by the function and by the mode of application. Protective functions include: creating awareness of hazards, detecting and warning of abnormal conditions, recovering the system back to its normal state and protecting people and the environment from injury and damage, etc. The mode of application may be: technical safety devices, policies, standards, procedures, supervision, training, briefing, drills and personal protective equipment, etc.

Reason pointed out that organizations have a multitude of latent failures at any point of time and that the situation with these latent problems is very dynamic: they transform in time, some disappear, and some new ones are born (Maurino et al. 1995, p. 24). On the other hand, the existence of many latent failures alone does not necessarily lead to an accident. Organizations with major safety shortcomings can operate for years without accidents partly thanks to the multiple defences in depth (Reason 1990, p.197).

The Reason model leads to a safety management paradigm: safety is about detecting and fixing the latent failures before their negative potential materializes. Obviously, comparing the model with the current understanding of complex systems highlights its simplicity. Accident causation is presented in a sequential manner and there is no room for continuous rich interaction between the different artifacts, agents and barriers within the system. At the organizational level, the roles and contributions of the various levels, e.g. the top management, are quite simplistic and linear. For example, one is left with the question what really drives the management to act in a specific way, e.g. what the real purposes in the system are.

In the context of organizational decision-making, James Reason highlighted the asymmetry between safety and production in terms of goals and feedback. Decision-makers need to balance organizational resources between production goals and safety goals. Investments in production have relatively certain and measurable outcomes whereas safety investments are much more difficult to measure in terms of their outcomes. According to Reason, most of the feedback related to investments in production is positive: e.g. more marketing leads to more sales. It is also presentable in numbers and continuous. In contrast, feedback about safety arrives in the form of bad news when an incident takes place. Such feedback is not continuous and it is difficult to get meaningful numerical data on safety. All these factors, according to Reason (1990, p. 204), bias the decision-makers to put more weight on the production goals than they perhaps should, relative to the safety goals.

1.3 Safety management systems, SPI's and bureaucratization of safety

One of the significant trends affecting safety management within transport has been the introduction of *safety management systems* (SMS) within several domains roughly during the last 15 years. The international civil aviation organization (ICAO) rolled out its safety management system concept in the first years of the millennium with applicability dates for the new standard ranging from 2001 to 2013 (ICAO 2017). In addition to the standard, ICAO produced its own safety management system framework and started running courses around the world to educate civil aviation authorities on the topic. The ICAO

Safety Management Manual was initially published in 2006 and had its third edition in 2013 (ICAO 2013). Safety management systems have been established also in the railway domain (see e.g. ERA 2015a) and even proposed for private transport companies, e.g. operating trucks (see e.g. QTA 2009).

The various safety management system frameworks describe in detail all the policies and processes that are deemed necessary for ensuring proper safety management at the level of a single operator, e.g. an airline or a train operator. As the topic of the dissertation is risk management at the level of the transport system, the detailed content for safety management at the operator level falls outside the scope. However, there are a few relevant aspects worth focusing on.

The first one is the role of risk management within the safety management system. Risk management is an integral part of SMS's and is typically split in different types of risk assessment tasks within the framework. For example, the ERA "SMS Wheel" includes a segment on risk assessment (ERA 2015b), covering (in simplified language):

- Control of risks associated with the activity of the train operators and railway track maintenance/works
- Risks arising from the activities of other parties
- Change management (risk evaluations related to changes; including human factors)
- Compliance with legislation, rules & standards
- Coordination tasks for track management

Similarly, the ICAO SMS framework includes (ICAO 2013, pp.160-161):

- Safety risk management, covering:
    - Hazard identification
    - Safety risk assessment and mitigation
- The management of change (which typically leads to risk analyses, similarly to ERA SMS)

The ICAO framework has dedicated content for the national civil aviation authorities - this part is called the State Safety Program (SSP). Its content in terms of risk management is limited: the basic idea is to require the operators to carry out proper risk management while the authority is also encouraged to promote risk approaches in its own work (ICAO 2013, p. 71).

The second interesting point is that both for ERA and ICAO, the risk management process is fed with "safety data" through hazard identification. This together with the promotion of Safety Performance Indicators (SPIs) reveals the desire to lean on collected operational data. This tendency also hints that such data is available, and there is willingness to "put it to work". As discussed above in Chapter 4.1.1, such data is classically reflecting the "negatives", e.g. undesirable safety events. For an example of potential event data (and descriptive factors) from the maritime domain, see Ladan & Hänninen (2012), and for aviation examples, see ICAO (2013, p. 35). Especially in aviation, the safety data collection methods have evolved continually leading to new standardized processes such as Flight Data Analysis (FDA) and live observation of flight crews on real commercial flights (e.g. through the Line Operations Safety Audit – LOSA – method). For an illustrative example of the volume and potential uses of such data, see the Evidence Based Training data report (IATA 2014).

The third point of interest is the question on how safety as a priority is seen in relation to other important organizational priorities: are other priorities (e.g. financial performance) ignored, are there implicit assumptions about them, or are they addressed openly, taking a holistic view to all the priorities? The ICAO SMM (2013, p. 29) does acknowledge that there needs to be a balancing act between production goals and safety goals and uses the metaphor of "safety space" to illustrate this. However, there is no discussion about other priorities and the practical guidance does not include tangible methods in integrating different types of organizational risks or dealing with conflictual interests in decision-making. The ERA SMS guidance (2015b) is very similar: it focuses on safety and safety management, and for example when "asset management" is discussed it is seen simply as an enlargement of safety management process to include important assets of the company.

Finally, one can ask whether there are safety-related problems that these SMS frameworks do not seem to be able to address. Referring back to the discussion on complex adaptive systems, at least two weaknesses can be pointed out in the SMS approach. First, there seems to be a somewhat idealistic notion about the use of power in the organization and about giving safety a very high priority, continuously. Numerous discussions that the author has had with people responsible for safety management in airlines over the last 15 years make it very clear that the top management has an almost total power on what can be done in safety management domain, and in the case where the top management would not support the safety work, the whole SMS would be severely handicapped. Secondly, the SMS approach falls short of proposing a solution against the "drift into failure" and the related mechanism of normalization of deviance. As Dekker (2014) points out, the daily frustrations, workarounds and improvisations are considered so normal that they do not typically rise to the level of report-worthy incidents. Therefore, they may remain invisible to the safety management system.

In parallel to the introduction of safety management systems there has been another clear trend in safety management which is the desire to introduce quantitative safety indicators - so-called Safety Performance Indicators (SPIs). These are typically a part of the "safety assurance" function within the SMS. For example, the ICAO and ERA SMS frameworks (and even the QTA framework) include quantitative safety performance indicators (ICAO 2010). Based on the discussion in the previous chapter about safety as an emergent property within a complex adaptive system it should be obvious that trying to reflect the "safety level" of an organization in a quantitative manner thought SPIs is a huge simplification and may produce quite misleading data. Nisula (2014a) argues that the most important pragmatic limitation of safety indicators (which are typically based on *outcomes*) is that they simply count *how many times* certain kinds of events occur and remain completely insensitive to the *content* of the event. Based on the role of events related to complex adaptive systems (see Chapter 3.5.4) and the questionable correlation between incidents and safety (see Chapter 4.1.1), the use of safety events as a measure of safety could be put in question even more fundamentally.

As a related trend to the SPI approach, Dekker (2014) argues that safety management has become an increasingly bureaucratic activity. He argues that safety management has partly shifted away from the true experts (e.g. workers with their tacit knowledge) who know the operational environment, and is being carried out by "bureaucrats" who are far away from the real activity. The safety numbers (like SPIs) have become the standard currency to prove safety performance within and between organizations and safety accountability upwards has become more and more important compared to safety responsibility towards the operation. Especially when coupled with incentive schemes, such approaches can lead to "numbers games" where "looking good" becomes more important than true concern over safety risks, and even to suppression of evidence. Consequently, risks may seem to be better under control than is actually the case.

2 Safety II: safety as resilience

Interestingly, James Reason had created the concept of *safety health* already in the 1990's (Maurino et al. 1995, p. 159). According to Reason, an organization would try to move within the *safety space* to the zone of maximum possible *resistance* and stay there. For Reason this was the only realistic goal of safety management. He also commented that due to less and less negative feedback and the resulting complacency, *staying* in the good zone would be more challenging than first getting there. Based on these concepts, Reason developed methods in order to try to assess the safety health of an organization by sampling details which could point to problems in the workplace and reveal organizational factors behind them (Maurino et al. 1995, pp. 161-165). It is easy to see the similarity between Reason's term *resistance* and the *resilience* concept. Moreover, the safety health concept was remarkable in the sense that different organizations who were not experiencing accidents could still be considered to be at different levels of safety health, i.e. have different levels of safety margins.

In terms of its Latin origins, the term *resilience* has the meaning to *rebound*, to spring back. There are various scientific definitions for resilience in different disciplines. According to Laprie (2008), a

common point to the definitions is to see resilience as the ability to successfully accommodate unforeseen environmental perturbations or disturbances. Resilience is not directly related to the stability of the system itself, i.e. a system which has low structural stability could still be very resilient. Typical terms/expressions associated with resilience include (Leviäkangas & Aapaoja 2015):

- capacity to absorb disturbances and shocks
- ability to maintain the capability to function
- recover to the original state
- prepare for and adapt to changing conditions
- withstand and recover rapidly from disruptions
- resist, absorb, accommodate and recover
- ability to maintain critical service level

Wreathall (2006) offers the following definition for resilience:

> "Resilience is the ability of an organization (system) to keep, or recover quickly to, a stable state, allowing it to continue operations during and after a major mishap or in the presence of continuous significant stresses."

Sometimes, resilience has been defined as the inverse of vulnerability, thus including *exposure* (to relevant hazards) as one of the factors (e.g. Leviäkangas & Aapaoja 2015). On the other hand, for example for Aven (2011b, p.517), vulnerability and resilience do not incorporate the uncertainty or probability dimensions.

Woods (2015) points out that the concept of system resilience has become hyper-popular over the last few years and consequently there are many different proposed meanings and definitions for the concept. He organizes the different views around four basic concepts:

- Resilience [1] as *rebound*. According to Woods, the ability to recover does not depend on what happens *after* a surprise but on the capacities present *before* the surprise that can be deployed or mobilized to deal with the surprise. A surprising event can be considered a brutal and abrupt audit: the focus should not be only on the attributes of the event itself but also on how it challenges the model behind the capabilities built in the system. Woods also points out that the concept of recovery to normal or previous function departs from reality in that the process of adapting to disruptions and surprises over time changes the system in multiple ways and even when the goal is to return to the previous state, the adaptation changes both the system and its environment.
- Resilience [2] as *robustness*. Robustness refers to an increased ability to absorb perturbations. According to Woods, robustness is a relevant concept only for disturbances which are well modeled, while the real interest of resilience is in cases where disturbances are not well modeled and introduce a surprise. A system may be robust within the set of modeled disturbances but remain brittle at its boundaries when an event outside the set is experienced. He also argues that there is a fundamental trade-off for complex adaptive systems where expanding the system's ability to handle some additional perturbations increases the system's vulnerability for other types of events, in the face of disturbances that fall outside the considered set of disturbances.
- Resilience [3] as *graceful extensibility*. Graceful extensibility can be seen as the opposite of brittleness: extending adaptive capacity in the face of surprise. Irrespective of the system's performance when operating well within its boundaries, the focus is here in the system's performance near and beyond its boundary: does brittleness lead to a rapid decline in the system performance or can the system *stretch* to handle a challenging surprise? Woods uses the term graceful extensibility (instead of graceful degradation) because extensibility at boundaries can be very positive and lead to success - and not simply less negative capability.
- Resilience [4] as *sustained adaptability*. This view to resilience stems from the observation that some layered networks or complex adaptive systems demonstrate sustained adaptability but most layered networks do not. There are three questions to ask:

- Which characteristics explain that some networks produce sustained adaptability and others do not?
- Which design principles or techniques would be key in succeeding to produce a network with sustained adaptability?
- How could one verify whether a particular system has the ability for sustained adaptability over time?

Woods (2015) argues that the yield from the first two concepts about resilience (rebound and robustness) has been low and that focusing on the rebound aspect misdirects the inquiry to the reactive phases only. For him, the promising parts of resilience research focuses on resilience [3] and [4]: graceful extensibility and sustained adaptability.

2.1 Resilience detection and enhancement as a means for Risk Treatment

Celebrated examples of resilience in the form of improvised creative solutions in emergency situations include (Fairbanks et al. 2014):

- Using car batteries from nearby vehicles to power the control room instruments of the Fukushima nuclear power plant during the 2011 disaster.
- Apollo 13 recovery after the onboard explosion in 1970, impressively reconstructed in the 1995 film with the same name.
- Landing the United Airlines DC-10 in Sioux City in 1989 despite having lost all three hydraulic systems and consequently all flight controls of the aircraft.

It seems obvious that increasing resilience reduces risks, and vice versa.

In resilience engineering, success and failure are seen as outcomes of the same underlying behavior. Variability and unexpected events are considered natural parts of system operation (Rankin et al. 2014). Safety is created at the sharp end when practitioners interact with the hazardous processes, face the multiple demands and use the available tools and resources. They encounter difficulties, complexities, dynamics and trade-offs and are expected to fulfill several, often conflicting, goals (Woods & Cook 2002). To quote Fairbanks et al. (2014), such difficulties include:

> "…combinations of usual and unusual demands; environmental disruptions; variations in staffing or other resources; information losses or corruptions; diffuse, varying, or conflicted goals; and, critically, incessant change. It is the resilience of these systems that gives them the ability to produce success despite conditions that could easily lead to failure."

Woods and Cook (2002) underline the role of sharp end practitioners:

> "Ultimately, all efforts to improve safety will be translated into new demands, constraints, tools or resources that appear at the sharp end. Improving safety depends on investing in resources that support practitioners in meeting the demands and overcoming the inherent hazards in that setting."

Unfortunately, everyday resilience can easily go unnoticed. It is argued that to understand failure requires first understanding how practitioners usually achieve success in the face of complexities and threats (Woods & Cook 2002). According to Rankin et al. (2014):

> "In organizations today, critical details of how practitioners cope through everyday adaptations are often not recognized, documented, or acknowledged and are known only as implicit knowledge by individuals and teams."

This is quite shocking. Enhancing resilience is a promising way to reduce risks, and safe systems likely have a lot of inherent resilience, but the resilience is not identified and captured. Rankin et al. also point out that sometimes important functions might even be "designed away" because of this lack of

understanding. Hollnagel (2014, p. 40) uses specific terms Work As Done (WAD) and Work-As-Imagined (WAI) to highlight the difference between the two. The former refers to the real work in its real context with all the adaptations and challenges referred to above. The latter represents the bureaucratic or management view on how the work *should be done* according to prescribed policies and procedures - in an (at least somewhat ideal) normative situation. Thus, if WAD could be captured, potentially a lot could be learned both about the weaknesses and the inherent resilience of the system. This is relevant both from the point of view of risk identification and treatment.

Finding out the sharp end practices that actually create resilience in an existing organization is a challenge in itself. Woods & Cook (2002) highlight that on one hand the distant views of observers cannot capture the actual experience of those who perform technical work in context, and on the other hand the practitioners' own descriptions of their work are often biased and cannot be taken at face value. Despite the challenges, the approach of resilience engineering is to try to understand how success is created in the everyday operation and how practitioners cope with variations that fall outside of the organization's formal instructions and procedures. Another interesting perspective is how the system itself learns about safety and responds to threats and opportunities (Woods & Cook 2002).

According to Woods (2015), both surprise and the way in which adaptive systems succeed and fail have regular characteristics. In terms of the predictable challenges over the life cycle of a system, he lists the following:

- Assumptions and boundary conditions will be challenged, i.e. there will be surprises.
- Conditions and contexts of use will change.
- There will be adaptive shortfalls and some people will have to fill the breach.
- The factors that produce or erode graceful extensibility will change multiple times.
- The system will have to adapt to seize opportunities and respond to challenges by readjusting itself and its relationships in the layered network.

In terms of the regular characteristics of failure, Woods states that the starting point for failure is exhausting the capacity to deploy responses as disturbances grow: this pattern is called *decompensation*. A related concept is *critical slowing down* where an increasing *delay* in recovery following disruption hints of a nearing saturation of its range of adaptive behavior and therefore an approaching collapse or a tipping point. Calling on resources to stretch repeatedly may overwork the system's adaptive capacity and result in consequences associated with *stress*. On the other hand, systems with high graceful extensibility are capable of *anticipating* bottlenecks and difficulties ahead. Observing how systems have been able to adapt to disrupting events in the past provides hints on their potential for adaptive action in the future. As mentioned above, if a system is to exhibit resilient behavior, the adaptive capabilities must be in place *before* the disruption arrives.

Fairbanks et al. (2014) state that the lack of requisite imagination about the range and nature of possible disturbances is concerning. Like Woods (2003) puts it: "The past seems incredible, the future implausible". This is one reason why *diversity* is often proposed as one of the key elements in creating resilience (see e.g. Laprie 2008 and Dekker 2011, pp. 173-176).

According to Fairbanks et al. (2014), resilience engineering is the deliberate design and construction of systems that have the capacity of resilience – a definition which resonates well with Woods' resilience [4]. A holistic view to resilience needs to be adopted as the challenge will be the management of adaptive capacity among multiple conflicting interests. Trade-offs are unavoidable as resilience in one area (e.g. maintaining service provision) often compromises resilience in another area (e.g. safety) as Johnson & Shea (2007) point out. Woods (2015) proposes that it makes sense to judge whether a system is resilient based on how well it *balances* all the trade-offs. Sometimes the situation is further complicated if the people who need the resilience and the people who have to pay for it are not the same (Leviäkangas & Aapaoja 2015). As resilience is not easily converted to numbers, it may easily receive less attention than phenomena that are easily measurable and consequently is under constant risk of being lost in the midst of actions optimizing economic returns (Fairbanks et al. 2014). Also, like Woods and Cook (2002) point

out, due to performance pressures from stakeholders, the benefits of change are often used to increase productivity and efficiency rather than to make the organization more resilient.

Another approach to organizational resilience can be taken through the so-called *high reliability organizations* (HRO). This branch of research has tried to capture observed commonalities of operations within organizations which exhibit effective management of innately risky technologies through organizational control of both hazard and probability (Sutcliffe 2011). The HRO work is focused on organizations which operate in unforgiving social and political environments, operate risky technologies and for which the scale of possible consequences from errors or mistakes precludes learning through experimentation. These organizations use complex processes to manage complex technologies. In practice, the domains under study have been aircraft carriers, air traffic control and nuclear power. From this research five organizing principles have emerged:

- *Preoccupation with failure*: being chronically aware of the risks and potential hazardous events.
- *Reluctance to simplify interpretations*: encouraging different opinions and seeking alternative perspectives so that assumptions can be questioned and a more complete picture of the current situation can be created.
- *Sensitivity to operations*: ongoing interaction and information sharing at the operational level to maintain an integrated big picture of ongoing situations so that necessary adjustments can be made and people learn about each other's talents and skills.
- *Commitment to resilience*: developing capabilities to cope with mishaps, learning about them and anticipating future problems.
- *Deference to expertise*: during high tempo situations like during a crisis, decision making migrates to the people with the most relevant expertise, irrespective of authority or rank. People know the expertise areas of the others.

Despite the challenges, there are tools and methods for working on resilience. Hollnagel (2011) has proposed a so-called Resilience Analysis Grid (RAG). In short, Hollnagel identifies four key abilities that together constitute resilience:

- Ability to respond
- Ability to monitor
- Ability to anticipate
- Ability to learn

In the Resilience Analysis grid Hollnagel introduces specific probing questions for each of the four abilities in order to help evaluating the level of resilience. Rankin et al. (2014) introduce a framework for capturing and analyzing adaptations in everyday work situations. Bergström et al. (2009, 2011) report on experiments where the training of flight crews and maritime crews was modified to focus on the development of generic competencies in addition to focusing on the operating procedures, in order to build resilience for escalating situations. The concept of evidence-based-training (EBT) in aviation follows the same line of thought (see e.g. IATA 2013, pp. 5-6). In 2013, Eurocontrol produced a publication called "From Safety-I to Safety-II: A white paper" (Eurocontrol 2013). The publication explains the transition from the safety-I paradigm to the safety-II paradigm in great detail. The so-called safety factors (Nisula 2014b, 2015b) offer a framework for capturing resilient practices in operations.

2.2 Contemporary safety challenges and paradoxes

Morel et al. (2008) argue that the practitioners who can perform with the best levels of resilience are the ones who are exposed to significant risks frequently and have therefore acquired the know-how how to survive the challenging conditions. Morel et al. point out that this leads to the paradox that some of the most resilient activities are also the ones where the greatest risks are taken. Exposure to dangers and frequent decision-making situations (with trade-offs between conflicting goals) create inherent "natural resilience" among the practitioners. Morel et al. argue that "it is extremely difficult to help an operator to acquire these skills without exposing him or her to risks". These observations underline that the relationship between resilience and safety is not simple.

As opposed to such natural resilience, most safety critical activities count on "safety through constraints" – such as prohibitions and protections. Morel et al. (2008) postulate that:

Observed safety = $[S_C + S_M]$ where S_C denotes Constrained safety
and S_M denotes Managed safety (or resilience)

For example, safety on sea-fishing vessels was concluded to be almost entirely based on managed safety, through the autonomy of the fishing skippers. However, it was also concluded that this type of innate resilience within the craftsmanship system was unable in itself to provide high safety levels (Morel et al. 2008).

On the other hand, more safety could kill the existing fishing system, due to too rigid and expensive constraints. This introduces another paradox and reminds of the above-discussed fact that safety can never be discussed in a vacuum. It is only one of the strategic (and survival-critical) priorities. Morel et al. (2008) report that professional fishers do not ask for more regulations but for better means for staying at sea in unfavorable conditions. The ethical paradox for the state agencies then is whether they should support such production-optimizing efforts or deny them due to potential added risk-taking.

A critically important conclusion from Morel et al. is that *increase in the constrained safety almost always comes with a decrease in the managed safety* and the resilient adaptive ability of the system, which becomes more rigid. However, they state that the compatibility between the constrained and self-managed safety remains an open research question.

Amalberti (2001) discusses the paradoxes of almost totally safe transportation systems. He focuses on systems which reach the safety level of one disastrous accident per 10 million operations and calls these systems *ultra-safe*. He offers scheduled airline operations and European railway operations as examples of such systems. Amalberti presents several characteristics that can be associated with ultra-safe systems:

- The safety level of these systems becomes asymptotic close to 5×10^{-7} disastrous accidents per number of operations/movements/km (whatever is the relevant reference unit or "base rate"). Amalberti calls this safety level the "mythical frontier".
- Solutions designed to reach this safety level tend to have devious effects when trying to advance beyond it. For example, the effect of automation as a solution starts wearing out before the mythical frontier – probably due to weakening situation awareness.
- Current ultra-safe systems tend to be ageing, over-regulated, rigid and highly unadaptive.
- Accidents occur in the absence of any serious breakdown or error, as combinations of factors. Such combinations are difficult to detect and to recover with traditional safety analysis logic.
- Reporting becomes less relevant in predicting major disasters.
- Due to low accident (and incident) rates, the system produces little feedback and evidence on what the safety level is and what measures have had a positive/negative impact on safety. Therefore, safety may become a political rather than scientific subject, with short-term having preference over long term measures.
- Over-stretched performance can give rise to new risks and reduce safety margins and recovery opportunities in degraded conditions.

The role of (the decreasing number of) incidents in very safe systems is somewhat controversial: are incidents markers of resilience or brittleness? Woods and Cook (2006) ask exactly this question and offer the following answers:

- Incidents show how the system can stretch to meet disruptions, but also the limits on that adaptive capacity.
- Incidents reveal boundary conditions and how the system behaves near these boundaries.
- Incidents are therefore sources for evidence of otherwise hidden sources for adaptiveness and also illustrate limits.

- The system's response to disturbances is the basic unit of analysis for assessing resilience.
- Incidents need to be analyzed to understand when the challenges stress or even fall outside the system's competence envelope, how the system engages its adaptive capabilities and what are the limits.

In line with this, Amalberti (2001) argues that beyond a certain incident-reduction quota, the absence of incidents may actually increase the accident risk. He proposes that optimum safety is obtained by tolerating and monitoring a residual number of incidents, and that the total volume of failures may be more telling than the various details.

Amalberti (2001) postulates that today's systems may not be able to go beyond the mythical frontier of 10^{-7}. He argues that every system has its life cycle, during which it gets optimized and after which it gets replaced by a new system. The main operating principles of the system will determine the maximum safety level that can be reached with the system. The most suitable safety measures are always a function of the system's optimization level. Once the system has been highly optimized (or over-optimized, like Amalberti sees most of today's man-machine systems stemming from the 1960's), the objective is no longer to improve, but just to contain the system and to try to avoid "the" disastrous accident which could mark the end of the system.

In his 2013 book "Navigating safety" Amalberti goes further in his reasoning on systems' safety. He proposes the following:

- Risk management must treat all types of risk which threaten the survival of the enterprise/organization, not only safety related risks.
- Safety management should be part of a holistic total management approach, not to augment other types of risks when safety risks are reduced.
- One must address both the normal (optimized) operation and sufficient resilience in exceptional events. This goes back to the equation "safety = regulated safety + managed safety". In ultra-safe systems the latter should be addressed through specific actions.
- There are no known examples of systems which would have succeeded in maintaining expertise for exceptional situations in parallel with a procedural safety envelope.
- In real life, the hardest challenges are drift and violations, which augment the safer the system becomes.
- Working on the so-called weak signals is a seductive but fallible approach. A specific method would be required, as classic risk management systematically ignores weak signals. However, there is an infinite number of signals, so the time, resource and expertise requirements would explode.
- The organization should also be prepared to survive an accident.
- Safety becomes a concept of social perception and a kind of civilian right. But social forces are often irrational, e.g. pushing more funds to crisis management than prevention.

According to Amalberti (2013, pp.76-77), the difference between ultra-safe and not so safe systems comes from:

- Performance-limiting boundaries from an Agency.
- Total system view instead of only individual concerns.
- Standardization instead of artisan approach.
- Overprotection through procedures etc.

Paradoxically, in ultra-safe systems, nobody can really know what the real safety level is. One consequence is that safety communication loses its rationality. Nobody also knows which elements and factors contribute to high safety and which don't. Consequently, the system cannot clean itself anymore and becomes more and more constrained.

According to Amalberti (2013, pp. 54-55), the four main limits of the Reason model are:

- Linearity

- Cartesian view promoting reductionism. This approach is weak on accidents without clear failures.
- Idealistic view promoting the elimination of all latent failures. In today's world, the realistic assumption is to accept some "holes".
- Ideal of eliminating all accidents and incidents. Going to an ultra-safe system often means that natural resilience is lost and remaining accidents are devastating.

The last point highlights two related paradoxes which are perhaps the greatest paradoxes of safe systems:

- The safest systems may produce the most devastating accidents.
- The safer the system, the less tolerable are the remaining accidents for the public.

These reviewed challenges and paradoxes need to be kept in mind when risk treatment strategies and interventions are designed, even if the task becomes even more difficult.

3 Human Factors in production and decision making

This section discusses aspects of human performance and human behavior which can be broadly placed under the title of human factors. First, a very brief summary is made of the human factors research in relation to safety. The objective is to highlight the key findings in this field, significant for understanding the modern approach to safety and the implications on risk treatment. Secondly, some of the key human biases are discussed in the context of risk perception and decision-making. The objective is to show that the way humans work on risks is not a simple machine-like mathematical process. The final aim of the presented material is to be able to take into account as much of it as possible in the design of the risk management framework, so that the related methods are as well adapted as possible to the way human cognition works.

3.1 Human factors of operations and production

This section summarizes the key findings of human factors research applicable to safety and risk management. The title reflects the fact that most of the human factors topics discussed here apply to frontline operators, e.g. people like pilots, seafarers and drivers of vehicles, within the transport operations, i.e. in production.

As Amalberti (2001) points out, after reliability of technology had been improved through the reduction of technical breakdowns, it seemed as a viable concept to try to improve safety through the reduction of human error. This approach was explored for about 20 years and the results were summed up towards the end of the 1990s.

Amalberti summarizes the conclusions of this research effort in four points:

- Mistakes are cognitively useful and cannot be totally eliminated. *Error detection* and *recovery* are part of normal activity and often the true manifestation of expertise and more important than the actual production of errors/failures.
- The key objective is overall safety and not errors in themselves. Therefore, it is necessary to consider the entire contribution of individuals to systems operations. One way of doing this is through the Reason's model and another is the High Reliability Organization approach – both discussed above.
- Mistakes made by individuals are consequences of the systemic migration of sociotechnical systems towards augmented complexity, performance, and individual advantages. The normal operation may often be characterized with pressure to optimize production to highest feasible levels where deviance easily becomes the standard. Such operation with reduced safety margins at quasi-incident levels may result in brutal transitions towards loss of control and accidents.
- The subject of study is no longer the actual error-producing mechanism, but the study of the cognitive mechanism used to control risk. The human operator does not regulate the risk of error but tries to meet the performance objectives at the lowest possible execution cost, balancing

different factors such as the style used to control the activity (automatic/conscious), the way to protect against errors, mechanisms used to detect and recover errors, the way to remain aware of one's performance and tolerance of residual errors. Such factors can be regulated in real time during the activity. The resulting balance can be called natural safety or ecological safety: the objective is to control errors within an acceptable margin rather than suppressing them completely.

Furthermore, Amalberti (2001) presents the following points based on studies on error control:

- Human operators produce more or less a constant rate of errors whatever their expertise, except for absolute beginners. The average is one to three errors per hour. The number of errors tends to *decrease* in *more demanding situations* due to increased cognitive control, but this comes at the price of a collapsing recovery rate (all resource used by online control).
- Routine based errors increase with expertise whereas knowledge-based errors decrease.
- The flow of errors stays under control: 75% to 85% of errors are detected. Routine based errors have a higher detection rate than knowledge-based errors. Experts tend to disregard an increasing number of errors which have no consequence on the final objective of the work: 28% of non-recovered errors compared to 8% for beginners. This shows that for humans the reference is not the number of errors as such but the feeling on whether the situation remains under control.
- When the human operator starts moving out from the "field of safe operations" and from the feeling of being in control, she/he comes to a turbulent area and receives cognitive alarm signals, warning about the approaching loss of control. Making sure the first signs of losing control are very apparent is a logical target for safety measures. Detected errors contribute to increased situation awareness. Unfortunately, automation sometimes works in the opposite direction introducing overconfidence.

Another illustrative summary of the role of human error in safety was presented in the book *Behind Human Error: Cognitive systems, computers, and hindsight* written in 1994 by four scientists at the Ohio State University (Woods et al. 1994). The authors first highlight the problem of explaining accidents simply with the label "human error" – for example (pp. 1-2):

> *Surveys of anesthetic incidents in the operating room have attributed between 70 and 75% of the incidents surveyed to the human element. Similar incident surveys in aviation have attributed over 70% of incidents to crew error.*
>
> *The typical belief is that the human element is separate from the system in question and, hence, that problems reside either the human side or in the engineered side of the equation. Incidents attributed to human error then become indicators that the human element is unreliable.*

They then present premises for research on human error (pp. 11-30):

1. Errors are heterogeneous.
2. Erroneous actions and assessments should be taken as the starting point for an investigation, not an ending.
3. Erroneous actions and assessments are a symptom, not a cause.
4. There is a loose coupling between process and outcome.
5. Knowledge of outcome (hindsight) biases judgments about process.
6. Incidents evolve through the conjunction of several failures/factors.
7. Some of the contributing factors to incidents are latent in the system.
8. The same factors govern the expression of expertise and of error.
9. Lawful factors govern the types of erroneous actions or assessments to be expected.
10. Erroneous actions and assessments are context-conditioned.
11. Error tolerance, error detection, and error recovery are as important as error prevention.
12. *Systems* fail
13. The design of artifacts affects the potential for erroneous actions and paths towards disaster.

In conclusion of the above, on one hand, the concept of human error is still valid and the understanding of the mechanisms behind human error has greatly enhanced in the last 30 years. Good principles in design for ergonomics and working practices are important in order to minimize all avoidable errors. On the other hand, however, what finally matters is successful task completion, and for this error detection, recovery and tolerance are as important as the emergence of errors in the first place. Even in good operations, plenty of errors take place every day without necessarily putting the system in danger: an expert making errors with no negative impact is a very different situation from someone making mistakes and thereby not reaching the desired outcome. The observation that not all errors contribute to negative outcomes highlights the important fact that *accidents (in safe systems) cannot be explained simply as human errors*. A more systemic view also shows that human errors are influenced by many system characteristics (e.g. complexity, production pressure) like Amalberti's conclusions above point out. As a summary, it can be stated that on one hand *error counting is not a good measure for the safety of the system, and on the other hand, focusing only on error prevention is not usually the ideal way to enhance system safety*. From the risk management point of view, overall safety (at the system level) and maintaining control (at the human operator level) are the key objectives, and safety interventions should focus on these. For example, for an expert user, making the first signs of losing control visible would probably be a better intervention than trying to prevent all errors.

3.2 Human biases, risk perception and decision making

Unfortunately, besides the somewhat impressive capabilities of people, humans are also subject to multiple cognitive biases and limitations in analyzing information and making sensible decisions. These have direct consequences on the capability to run an effective risk management process. The first set of limitations has to do with mental models and heuristics, especially in relation with complexity:

Imperfect mental models. Sterman (1994) reminds that all decisions are based on models. Mental models are the implicit causal maps that we hold concerning the systems around us. They contain our beliefs about the network of causes and effects that describe how the systems operate, the boundary of the model and the time horizon we consider relevant. Sterman argues that most people do not appreciate the ubiquity and invisibility of mental models and instead believe naïvely that their senses reveal the world as it is. Importantly, the mental models also guide what information from the environment is acquired and how it is interpreted. Human senses and information systems select only a tiny fraction of the possible experience and some of this filtering is based on (conscious or unconscious) decisions. Our past experience and expectations determine powerfully what we perceive in the environment. In this way, past perceptions and expectations limit learning and changing mental models as the anomalies that might challenge current models may be filtered out. Perrow (1984, p. 318) also argues that selecting the context before interpreting a situation is crucial but an almost effortless pre-decision act, made without reflection based on past experience. He argues that sometimes the human operators unconsciously pick what seems to be the most familiar context but not necessarily the correct one, setting the scene for bad decisions.

Flawed cognitive maps of causal relations. As a specific example of imperfect mental models, Sterman (1994) highlights that the cognitive maps humans hold are vastly simplified compared to the complexity of the systems themselves. Furthermore, people are unable to infer the dynamics of all but the simplest causal maps. This is a problem which becomes very significant with complex systems. The heuristics humans use to map causal relationships lead systematically to cognitive maps which ignore feedbacks, multiple interconnections, nonlinearities, time delays and other elements of dynamic complexity. People generally adopt an event-based, open-loop view of causality and in the reporting of information do not understand stocks and flows.

Inability to simulate mentally systems with feedback loops. Not only are humans unable to construct useful cognitive maps of complex systems, but they are also incapable of mentally simulating the dynamics of a complex system with feedback loops - even if they were given the correct representation of the system (Sterman 1994). Understanding the dynamics of multiple feedback loops requires intuitive solutions of high order differential equations – a task which exceeds the cognitive capabilities (except

perhaps in the very simplest cases). A typical example of this limitation is the fact that people significantly underestimate exponential growth.

Other human biases and human limitations mentioned by Sterman (1994) include:

- The *Fundamental Attribution Error*: this is the human tendency to attribute the behavior of others to dispositional rather than situational factors. The unfortunate consequence is that behaviors are seen as stemming from personalities and the contribution of system structure is missed. This diverts attention away from the systemic high-leverage points.
- *Overconfidence*. Humans tend to underestimate uncertainty, be overconfident of positive expectations while ignoring or downgrading negative information (Makridakis & Taleb 2009), and maintain an illusion of control, even on influencing the outcome of random events. Makridakis & Taleb (2009) highlight that forecasts done by experts are not more accurate than those of other knowledgeable people. Moreover, experts are less likely to correct their forecasts than nonexperts when new evidence emerges, disproving their earlier beliefs.
- *Confirmation bias*. Humans tend to seek evidence consistent with current beliefs rather than potential disconfirmation. For a detailed discussion of confirmation bias, see e.g. Nickerson (1998). According to him, evidence supports the view that no matter how open-minded one has been in forming a position on an issue, once that position has been established, one's primary purpose becomes that of defending or justifying that position. Confirmation bias may couple with *availability heuristic*: the importance and relevance of information is weighted based on how readily it is available (e.g. in one's own memory, see Tversky & Kahneman 1973).
- *Defensive behavior* is about avoiding public testing of important hypotheses and could be also seen as a kind of group-level acting of confirmation bias.
- *Groupthink*. Defensive routines at the level of a group lead to what can be called groupthink. In groupthink, the members of the group mutually reinforce their current beliefs, suppress dissent, and isolate themselves away from those with differing views and challenging evidence. The mental models of the team members may in this way remain hidden, ill-formed and ambiguous. Waring (2015) argues that groupthink frequently builds around an authority figure and that groupthink can create blind spots that affect both risk perception and responses.
- *Framing issues*. Human judgment may be strongly affected by the frame in which the information is presented, even if objective information is unchanged. Perrow (1984, p. 318) discusses framing issues in the context of the correct use of base rates in probabilistic reasoning. He argues that depending on how the problem is presented, the subject will or will not use the base rates and probabilistic reasoning.

Interestingly, De Martino et al. (2006) studied the framing bias from the physiological perspective (in terms of brain activity) and concluded that the framing effect was associated with amygdala activity, suggesting that *the emotional system mediates decision biases*. They write:

> "…individuals incorporate a potentially broad range of additional emotional information into the decision process. In evolutionary terms, this mechanism may confer a strong advantage, because such contextual cues may carry useful, if not critical, information. Neglecting such information may ignore the subtle social cues that communicate elements of (possibly unconscious) knowledge that allow optimal decisions to be made in a variety of environments. However, in modern society, which contains many symbolic artifacts and where optimal decision-making often requires skills of abstraction and decontextualization, such mechanisms may render human choices irrational."

There are many other known human biases related to decision making, in addition to the ones mentioned here. Like De Martino et al. (2006) state: "when taking decisions under conditions when available information is incomplete or overly complex, subjects rely on a number of simplifying heuristics, or efficient rules of thumb, rather than extensive algorithmic processing". However, as mentioned above, the intent here is not to be exhaustive but to review a sample of evidence to highlight the fact that human decision making is not necessarily a rational analytic process but something far more complicated and fragile.

Not surprisingly in the light of the mentioned limitations, human performance in complex dynamic environments is far from optimal and often even far from reasonable. It is very hard for humans to manage the typical unanticipated side effects and delayed consequences. Experimental studies confirm that human performance in such environments is poor not only relative to normative standards, but even compared to simple decision rules, e.g. "do nothing" (Sterman 1994). In different types of experiments, subjects created average costs which were 10 times greater than optimal and/or costly escalations in constant conditions, people were insensitive to time delays in the system, they may have missed the primary target (e.g. letting the headquarters to burn down in a forest fire simulation or letting patients sicken and die in a medical setting), and often subjects would blame their poor performance on external events. The greater the dynamic complexity, the worse people performed relative to the potential. People also seem to be incapable of adapting their decision rules as the complexity of the task increases: the time taken to make decisions in higher levels of complexity is not significantly longer than with simpler situations – sometimes it even gets shorter (Sterman 1994).

Another level of the problem is that people have great challenges to *learn* in situations of dynamic complexity (Sterman 1994). For example, after 50 years of simulated experience with perfect immediate feedback, 83% of the subjects still performed worse than a naïve strategy which doesn't utilize the available feedback from the system or react to the behavior of the competitor. In complex systems available feedback is ambiguous and mixed together with many other variables subject to simultaneous changes. Being able to learn in this type of environment would require a disciplined scientific approach based on diverse evidence and active challenging of beliefs. Unfortunately, humans do not generally work this way (Sterman 1994).

The second set of human limitations considered here are associated with risk perception. The phenomena discussed below make the already difficult task of risk assessment even more difficult by introducing additional error and systematic biases.

Tendency to underestimate uncertainty. Mosleh et al. (1988) report that overconfidence is one of the known biases in judgmental estimates. People have the tendency to give overly narrow confidence intervals which reflect more uncertainty than is justified by their knowledge. Tickner & Kriebel (2006) propose that one factor contributing to this tendency may be that admitting honestly the uncertainties that even an experts or decision-makers may face could weaken agency authority by creating an image of the agency as "unknowledgeable".

(In)tolerance of uncertainty. Johnson & Slovic (1995) discuss the tolerance of the public towards an openly communicated uncertainty in risk assessment. They report the following findings:

- The public is not familiar with uncertainty in risk assessment nor in science in general
- When presented simply, people may recognize uncertainty.
- Graphics do not necessarily help in communicating on uncertainty
- Trust in government and authority is an important factor in perceived risk and probably more important than openness on uncertainty.
- Agency's discussion of uncertainty on risks appears to signal honesty – but also incompetence for some people.

Ignoring very negative scenarios. The famous sentence of David D. Woods on some of the future scenarios being *implausible* was already mentioned above in the context of resilience (Woods 2003), as well as the alarming lack of requisite imagination about the range and nature of possible disturbances (Fairbanks et al. 2014). Makridakis et al. (2009) add that people don't fully face the possibility of a rare event occurring and its consequences. These observations fit well with the *overconfidence* mentioned above.

Biases in risk perception. Dionne et al. (2007) state that there is empirical evidence that perceived risk is generally biased and that perception of risk influences behavior. Perrow (1984, pp. 326-328) reports

on studies focusing on what kind of factors increase people's risk perception. The most important factor which was named "dread risk" was associated with:

- lack of control over the activity
- fatal consequences in case of mishaps
- high catastrophic potential
- reactions of dread
- inequitable distribution of risks and benefits (including the transfer of risks to future generations)
- the belief that risks are increasing and not easily reducible

The presence of these factors turned out to be the best predictor of higher perceived risk. The next most important factor for high risk perception was the presence of unknown elements, i.e. risks that where unknown, unobservable, new and/or delayed in their manifestation.

Sensitivity of risk aversion. Studies suggest that attitudes to risk aversion may be sensitive to many factors. Balsa et al. (2015) concludes that attitudes to risk aversion are influenced by the environment and more particularly by peers. Wilson et al. (2008) suggest that risk aversion works when it is done in relation to oneself or in relation to people known to the subject, but not for stranger-to-self. Kahneman & Tversky (1979) argue that risk aversion can even change for the same person in a relatively short time frame: a failure to adapt to losses or to attain an expected gain induces risk seeking. So a person who has not "made peace with his losses" could accept gambles which would otherwise be unacceptable to him.

Finally, one key bias needs to be mentioned, relevant both to risk assessment and to decision making as such:

Hindsight bias. Baruch Fischoff (1975) showed in his famous experiments that the fact of knowing the outcome of an event/story greatly influences the way the different elements within the story are interpreted. Importantly, reporting the outcome also produces an unjustified increase in its perceived predictability - it seems to have appeared more likely than it actually was. Initially, Fischoff called this phenomenon "creeping determinism". He reported that disclosing the outcome to subjects carrying out the experiments approximately doubled its perceived likelihood of occurrence. He also showed that the relevance attributed to any data elements within the stories was highly dependent on which outcome, if any, subjects believed to be true. In other words, people find the evidence to justify the outcome. Hindsight bias is a particularly unfortunate phenomenon in safety analysis because it makes analysts think (after the fact) that the outcome was more predictable and visible than it actually was in real time. This can easily reflect negatively on the sharp end human operators being part of the incident/accident. Due to increasing the perceived likelihood of the observed outcome, risk assessments on similar future scenarios get easily biased.

4 Synthesis and conclusions

The way to understand safety has evolved significantly in the last decades. The focus has shifted from individual sharp end practitioners and technical failures through Human Factors and organizational aspects to understanding safety as an emergent property in an (often) complex system. Safety management systems have been introduced and regulated within the transport operations. These frameworks are helpful in defining the set of elements that are deemed today's state-of-the-art practice and introducing risk management as a central activity within safety management. However, current SMS's still have several significant limitations. Safety is taken apart from other organizational priorities and the conflicts are not addressed. The Safety Performance Indicators offer a very simplistic approach to "measuring" safety and in the worst case can contribute to what Dekker calls the bureaucratization of safety. It is argued that current SMS's do not offer a solution for power issues (top management ruling against safety) nor for the symptomatic phenomenon of drift. Current SMS's are the products of the Safety-I era and despite the existence of Safety-II, the current practice at operators is still largely based on the Safety-I approach.

Resilience was defined and discussed at length. It is easy to see a natural link to the adaptability which is so vital within complex adaptive systems. Within resilience engineering, people are seen above all as the adaptive element, enabling a flexible and sustained operation, and no longer as the weak link in the system. From the operational point of view, one basic guideline for operators would be not to only optimize the normal operation but also be prepared for surprises, i.e. have resilience. This idea resonates well both with black swans and the natural variability in complex systems. On the other hand, this also reinforces the desirability to be able to identify the good resilient practices and to reinforce them, and ideally also get an idea of the overall organizational resilience. However, as discussed, finding the resilient features in a system can be challenging as even the practitioners themselves are not necessarily aware of the constant adaptation and the resilience dimension: all this is just "normal work" for them. The high reliability organization (HRO) approach was introduced as another way to approach resilience.

Current safety situation, especially within the very safe transport operations contains paradoxes. First, attaining the safest levels of operation has required a regulated approach but such proceduralization has a negative effect on the managed (or natural) resilience, which exists spontaneously in less safe systems. Secondly, safety is not only a blessing. Safety at the ultra-safe level can become so constraining and expensive that it threatens to kill the system, which may come to the end of its life-cycle and be replaced with a new system with new paradigms. Finally, the safest systems may also have the most devastating accidents.

In this dissertation, the aim is to embrace the Safety-II approach. The immediate consequence is that one should try to develop ways to measure safety on a positive scale, and not only as the absence of negatives. Additionally, it is also deemed vital to take a holistic approach to the organizations in question, so that safety is not treated separately from other organizational priorities and that safety risks are not presented and treated separately from other (corporate) risks. One aspect of taking the different priorities into account is that the safety and risk management functions themselves need to be carried out within acceptable levels of efficiency, resource and cost.

In most current transport operations, there is a strong tradition for safety data/information collection. As discussed, analyzing such data still has its place. Furthermore, with such a quantity of safety data available, it is safe to assume that a risk management framework which could not accept the data/information as input could hardly be acceptable to the transport safety community.

Various topics reviewed in this chapter point to the need to address safety and risks at several levels. The safety data typically highlights operational *events* (or incidents). The concept of resilience (including the high reliability organization approach) and the willingness to measure safety (on a positive scale) make most sense at the *organizational* level. The challenge of power "abuse" compromising safety, bureaucratization and drift require focusing at least on the organizational level but also on the *system* level, which goes beyond an individual operator. Similarly, the mentioned safety paradoxes imply the system level, which in this case could be a very large international system of systems.

The biases, heuristics and other human limitations discussed in the last part of this chapter stem from cognitive and even biological properties of humans and therefore are with us to stay. An approach which can be considered is to try to design various details within the risk management framework (and the associated processes and methods) in such a way that the biases could be counteracted as much as is feasible. Distorted risk perceptions and groupthink may be at least partially addressed by using diverse groups of people in teams where diversity and dissent is actively encouraged and the working methods aim at building knowledge around the topics to be assessed. To break free from common beliefs, people from outside the core group of analysts can be brought in – this should also help in enlarging perspectives in relation to black swans. The bad track record related to interacting with complex systems suggests that defining actions in a feedforward manner is not recommendable. Rather, an iterative approach based on adaptable experiments could be adopted. The impediments to learning within complex systems are

particularly worrying because they diminish hopes that the interactions could improve in time. Providing a solid theoretical background on complex systems is probably a vital basis. People can be encouraged to try to think at the system level, try to draft the possible interactions and feedback loops and compare draft models with real-life observations on system effects. However, realistically, learning will probably remain a considerable challenge. Concerning the role of uncertainly, the position taken here is that uncertainties should be admitted boldly and to enable this way of working without losing credibility with key stakeholders, it will probably be recommendable to run specific communications/educational efforts with stakeholders. In the case of a national transport safety agency, the stakeholders would include at least other agencies, the ministry of transport, people within the agency itself, and possibly the public.

Chapter 5 - Acceptability of Risks

An inevitable part of risk management is to find a way to define what are the limits of tolerable risk. The subject is treated in three sections. The first one examines the different ways in which the limits can be defined, the second section discusses the ethical base for risk acceptance criteria, and the third section reviews which existing criteria could be used or adapted to obtain the required method for defining the limits of acceptable risk. In practice, a key concern is to know what kind of safety investments are worthwhile compared to the associated risk reduction.

1 Risk tolerability as a multi-dimensional judgment

This section reviews the various contributions to risk tolerability in literature. Key points are deliberately formatted as bullets. The first three bullets come from the conclusions in Chapter 2.4.:

- It becomes clear that the *result of the risk assessment cannot be a simple number*. Presenting the results of risk assessment and taking decisions based on the results, becomes something more complicated and *multidimensional* than just comparing two numbers.

- There is value in trying to present the results in such a way that the *risk aversion policy is left open* and can be adjusted according to various needs.

- It is clear, that the proposed risk management framework will have to be able *to maintain awareness of potential black swans, irrespective of their perceived probabilities*.

Vanem (2012) states the following (formatting added):

- "*Different sectors will require different acceptance criteria*, and one cannot assume that it will be satisfactory to apply acceptance criteria from one sector to another one."

He also notes that different anchor points (for fixing the so-called FN-curves, see below) for societal risks have been suggested in the literature and in regulations, but that even for different ship types, these anchor points are proposed differently. As Amalberti (2001, 2013, p. 75) illustrates, different activities have very different safety levels and if these levels are used as a reference for risk acceptance (in line with the principle of *equivalency*; see Vanem 2012), the limits would indeed be very different between different activities. Moreover, Carlsson et al. (2004) report test result which indicate that subjects' Willingness To Pay (WTP) for a given risk reduction was significantly higher for flying than for travelling by taxi.

Similarly, according to Chilton et al. (2002), research in cognitive psychology and economics strongly suggests that preference-based Values for the Prevention of a statistical Fatality (VPF) may not be automatically transferable between different contexts. Their research generated estimates of preference-based values of safety of three domains (rail, domestic fires, fires in public places), *relative* to the value in road safety. This relative valuation was the preferred approach due to the low baseline levels of risk (e.g. for rail accidents) and the resulting high error margins if estimates of the absolute values were to be used.

Aven & Krohn (2014) argue that:

- In real life, risk cannot be measured in an objective way and the risk management needs to reflect this. Making judgments about the acceptability of risk on the basis of probabilities alone should be avoided.

Similarly, Aven & Ylönen (2016) state:

- Today, risk acceptance criteria are often used rather mechanistically by comparing calculated probabilities with predefined acceptance limits. Such use cannot in general be justified.

- Complex systems cannot be adequately risk analyzed using a static probability-based approach.
- Identifying the best solutions and measures cannot be determined from considerations of probability alone. Overall judgments are required.
- Such judgments must consider and make explicit: the supporting knowledge and related uncertainties; strength of this knowledge; assumptions; risks related to the assumptions; potential surprises relative to knowledge and beliefs.

Paté-Cornell (2002) states that the results of a risk analysis are generally meant to answer two kinds of questions:

1. Is a particular risk acceptable?
2. What measures can be adopted to maximize safety under resource constraints?

She points out that numerical values are only part of the answer to the first question: other aspects such as the controllability and the voluntariness of the risk must be considered. The general guideline for question two would be to support the optimal use of limited resources to provide the best protection to the maximum number of people. Furthermore, Paté-Cornell argues that:

- The search for an acceptable level of risk should focus on *an acceptable decision process*.
- Results of risk assessments can be compared only to the extent that they have been established on comparable bases (e.g. concerning all assumptions).

Aven & Vinnem (2005) and Ersdal & Aven (2008) propose some aspects for the risk acceptance decision process. They describe the method as multiattribute analysis with managerial review and judgment. Aven & Vinnem (2005) highlight as important elements the drive for generating *alternatives*, the ability to *communicate* the information from the analysis to the decision-makers (who need to understand it), and a good *visualization of the trade-offs* between safety and other aspects (such as cost).

2 Ethical basis for risk-related decision making

Vanem (2012) points out that there is an obvious link between the "right or wrong" of ethics and the "acceptable or not acceptable" risk judgments and the related criteria. The latter are important in protecting the public, environment and society as a whole against different risks, and according to Vanem it could be argued that the principles and the regulatory framework for risk management need to have an ethical foundation. Such a foundation should also be helpful in communication and justification of adopted risk acceptance criteria, for example towards the public. Vanem introduces the concepts of *values* (what is regarded as good), *moral norms* (how one should act), *moral agents* (a person who can have moral duties) and *moral subjects* (beings that should be taken into account in moral assessments). Perhaps the first ethical question is on whether only living human beings have a moral status or if it should be extended to animals, plants, the environment and future generations. Overall, Vanem reminds that there is no generally accepted ethical theory and the different theories may result in different judgments.

Abrahamsen et al. (2004) argue that life has a value in itself which cannot be measured in money. Risk reducing measures can be put in perspective by using the cost per expected saved lives (value of statistical life) but this is different from allocating a fixed value to life. Obviously, most people would not give up their lives against a certain amount of money. However, people may accept a certain level of risk against a certain amount of money. All this means that specific considerations have to be given to safety of human beings.

There are *individual* and *societal* risk acceptance criteria. Individual criteria are necessary when individuals or groups of individuals are exposed to specific (or additional) risks. On the other hand, for large systems exposing a large number of people to possible accidents, societal risk acceptance criteria are deemed to be the most appropriate (Vanem 2012). Societal criteria could typically be expressed in the form of the so-called FN-curves (linking accident frequency to fatalities) or risk matrices (linking the frequencies to fatalities in the form of discrete categories).

Vanem (2012) presents the following principles which could be used as the ethical basis for establishing risk acceptance criteria:

- *Absolute* risk criteria: an approach where a defined risk/safety level needs to be reached *irrespective of the associated cost*. An inverse approach would also belong under this principle: making risk decisions solely based on cost considerations.
- *As Low As Reasonably Practicable* (ALARP): here both risk level and cost are considered and risk reduction measures should be implemented as long as the cost is reasonable. The challenge is that some standard measures of practicality would be needed. See Jones-Lee & Aven (2011) for a detailed discussion on the interpretation of this principle.
- *Equivalency*: using existing activities or systems as a reference and requiring that an equal level of risk is obtained. This can be a good approach especially for novel activities or systems.
- *Maximum benefit to all*: here the principle is to try to maximize the total expected net benefit for the society as a whole. The challenge is to have an objective measure of the benefit. A widely used indicator is the expected length of life in good health for all members of the society. Problems can arise if there are risks which are not equally distributed among the population as some people might be sacrificed for the sake of others.
- *No individuals are to be sacrificed for the sake of others*: the idea is that certain individuals should not be exposed to additional risks in order to reduce the risk or increase the benefit to others. In such situations there should be a compensation to the most exposed individuals so that these "losers" are transformed into non-losers.
- *No mandatory risk reduction*: in this principle the idea is to let the economic system regulate itself autonomously based on the assumption that this is the most efficient way to manage the risks. This principle can generally be trusted only when the same stakeholder who is responsible for the risky activity also benefits from the risk reduction, i.e. it is unrealistic to assume that this approach will adequately address risks to third parties.
- *The accountability principle*: the decision-making process related to risks should be transparent and the decisions should be shown to be justifiable and communicable to the public. This full transparency shows clearly the political, societal and personal values underpinning the decision criteria. Quantitative criteria are preferred to qualitative, supporting an as "objective" assessment as possible.
- *The holistic principle*: safety decisions should be based on a holistic consideration of all risks across the complete range of hazards with all direct and indirect consequences. Failure to respect the holistic approach may cause unproportional expenditure of resource in some areas of the society at the expense of others.
- *Precautionary principle*. According to this principle, lack of solid scientific evidence on specific risks shall not be used to postpone cost-effective measures to address them. The related *cautionary principle* was introduced in Chapter 2.3. For more details on the precautionary principle, see Ahteensuu (2013) and for a discussion on different types of uncertainties in the context of the precautionary principle, see Aven (2011c).
- *Principle of parsimony*: This principle promotes simple and practical risk acceptance criteria over more complicated ones. This makes decision work easier and also supports clear risk communication to the public and to the stakeholders.

To these could be added two principles which are similar to equivalency (Hokstad et al. 2004):

- *Minimum Endogenous Mortality* (MEM). The current mortality rate (taking into account factors such as age and gender) should not increase significantly (e.g. by 5%) due to the introduced change.
- *Globalement Au Moins Aussi Bon* (GAMAB). The first word of this "globally at least as good" principle is important: it is the *combination of all residual risks* which should not exceed the current level after the introduced change.

Wolff (2005) adds the concept of *imposition of risk*: it is morally problematic when risks are imposed on people. The related societal concern is not based on a risk assessment but rather on moral criticism on those who are imposing the risk. For additional ethical topics, see e.g. Ersdal & Aven (2008).

It is good to be reminded of some of the prevailing expectations within the society. As mentioned previously, according to Amalberti (2013, p. 115), for systems which are considered very safe, safety becomes a concept of social perception and a kind of *civilian right*. Consequently, the accidents which occur – very rarely – are more and more intolerable for the public (p. 8). The public also demands greater transparency on risks but has *problems to deal with such transparency*: people are intolerant to all remaining problems within the safe systems and may even become doubtful that some information is hidden (p. 9). As accidents are very rare, factual information on what really creates the safety is not available, and for such systems *safety becomes a political rather than a scientific subject*, often leading to a short-term view (Amalberti 2001). Also, *social forces may act irrationally*, for example by channeling more funds to crisis management than prevention (Amalberti 2013, p. 115). The RSSB report "Railway safety and the ethics of the tolerability of risk" (Wolff 2005, p. 5) is an illustrative example of a situation where a very safe transport domain (railways in the UK) felt the need to justify its risk acceptance approach to the public.

Moving from the overall ethical principles to their implementation, Paté-Cornell (2002) reports that there is an emerging pattern on how such principles are reflected in risk tolerance decision making, across several fields and several countries. According to her, there are three categories:

- *Unacceptable risks* on an individual by someone else with probabilities in the order $10^{-4}…10^{-3}$. These are out of the range of cost-benefit analyses.
- *Tolerable risks* with probabilities smaller than the "de minimis" threshold with probability in the order of 10^{-7}.
- Between these limits of unacceptable and negligible, there is an area where costs and benefits generally enter the picture one way or the other.

Paté-Cornell uses the following limit values for annual individual risk (or risk per operation):

- De minimis for "workers" 10^{-6} and for public $10^{-6}…10^{-7}$.
- Intolerable for "workers" 10^{-3} and for public 10^{-4}.

However, she warns that in addition to cost/benefit considerations, risk acceptance also depends on factors such as voluntariness, controllability, familiarity, and (epistemic) uncertainty.

Similar to these three layers, Abrahamsen & Abrahamsen (2015) propose a layered approach for implementing the ALARP principle, where there is an "extreme economic perspective" (with decisions based on expected values), "extreme safety perspective" (based on cautionary principle) and a "somewhere between" area.

The adopted risk aversion strategy is another ethical consideration: are fatalities in a single catastrophic accident more important than fatalities in several small accidents? Chilton et al. (2002) report that in their study people considered deaths from different types of accidents (road, rail, domestic/public fires) equal (despite the potentially different number of deaths in road vs. rail accidents):

> "This therefore leads us to the conclusion that while people's priorities are indeed sensitive to the combined influence of the number of deaths, the psychological characteristics of hazards and social amplification effects following a major accident, in practice (at least using our particular elicitation methodology) it is the number of deaths which would appear to dominate the quantitative judgements people give."

Chilton et al. point out that there are two important caveats: first, not only the number of deaths should be considered but also the number and the quality of years of life gained. Secondly, the risk contexts in their study were quite similar and it may be that very different risks could lead to different results; for example, risks related to hazardous waste, genetic manipulation or food processing.

3 Methods and indicators

There are many different methods and indicators which are in use today to support decision making on risk acceptance. For example, Vanem (2012) discusses the following approaches:

- FN-diagrams. For detailed description how FN-diagrams are implemented in the maritime domain through the Formal Risk Assessment method, see IMO (2006). It is worth noting that the slope of the FN-curve indicates the risk aversion policy, so once the curve is established, the risk aversion policy is locked in.
- Gross Cost of Averting a Fatality (GCAF) and Net Cost of Averting a Fatality (NCAF) can be used with the FN-approach to justify meeting the ALARP principle.
- Quality Adjusted Life Year (QALE) and Healthy Life Year (HeaLY) can be used to take into account the degraded quality of life.
- Risk matrices (sometimes called "heat maps" due to the typical colour coding)
- Cost of Averting a Fatality (CAF, NCAF, GCAF) is used in assessing safety investments.
- Cost of Averting an oil spill (CATS) refers to the cost of averting a tonne of oil spilt.

Different reference values from different domains and countries for these and other indicators are available, in Viscusi & Aldy (2003), Jonkman et al. (2003) and Vanem (2012). Other indicators include:

- Value of a Statistical Life (VOSL), see e.g. Persson et al. (2001), Chilton et al. (2002) and Wren & Barrell (2010, p. 19).
- Value of Preventing a (statistical) Fatality (VPF), see e.g. Chilton et al. (2002).
- Willingness To Pay (WTP), see e.g. Jones-Lee & Loomes (1995) on the scale and context effects on WTP related to transport safety.

Reviewing the 11 bullets presented in section one of this chapter, it becomes obvious that the risk acceptance approach becomes a multidimensional question where no single indicator or criterion will alone suffice to make the decision. The existing practical methods and indicators become an important inspiration rather than something for direct application. Indeed, like Paté-Cornell (2002) states, the search for an acceptable level of risk becomes a challenge of *defining an acceptable decision process*.

Going away from mechanized risk acceptance procedures can be a challenge for everyone, including the agencies responsible for regulation and oversight. Like Aven (2013a) points out, the probability-based approach is attractive because compliance can be followed, controlled and supervised more easily than with a more holistic approach. He also offers the example of the use of the ALARP principle within the Norwegian oil & gas industry, reporting that the industry struggles with using it effectively, as it needs to be based on overall judgments instead of mechanized procedures.

Some guidelines exist *for relating different types of outcome severities*. The international maritime organization (IMO) has published guidelines for the Formal Safety Assessment, where the practice has been adopted that *one fatality equals 10 severe injuries and one severe injury equals 10 minor injuries* (IMO 2006, p. 7). Wren & Barrell (2010) present a comprehensive study of the total social and economic cost of injury in New Zealand. The European Aviation Safety Agency (EASA) follows a priority order in which people involved in aviation need to be considered in relation to severity assessment and accident prevention. The priority order is the following (EASA 2016, p. 17):

1. Uninvolved third parties
2. Fare-paying passengers in commercial air transport
3. Involved third parties (e.g. air show spectators, airport ground workers)
4. Aerial work participants / Air crew involved in aviation as workers
5. Passengers ("participants") on non-commercial flights
6. Private pilots on non-commercial flights

The background for this reasoning reflects both the level at which the people within the hierarchy are in control of the risks themselves, and pragmatic questions of resource/cost (European general aviation safety strategy working group 2012).

Another domain where risk acceptance criteria can be found is technical systems certification where the focus is on demonstrating that the probabilities of different levels of failures (and resulting assumed

outcomes) are tolerable. For example, the aircraft certification standards used by the Federal Aviation Administration (of the USA) require the probability of aircraft system failures which could lead to catastrophic consequences to be 10^{-9} for commuter and transport category aircraft (FAA 2011, p. 23, FAA 1998, p. 7). While these references can hardly be applied to the multi-dimensional risk acceptance judgments (e.g. including assumptions, uncertainties, strength of knowledge) they give an idea of the order of magnitude of required reliability levels. What can also be seen, is that for an aviation system generally at the safety level 10^{-6} (Amalberti 2001), the underlying technical system reliability requirements are three orders of magnitude higher.

The acceptance criteria for such low probabilities leads to the topic of black swans. It can be argued that managing risk acceptance judgments in a purely engineering context is fundamentally easier than doing it in an operational context. In any case, certain high safety industries are able to work with such low probabilities routinely. This also means that these considerations do not refer to black swans. It is worth noting that none of the listed numerical risk acceptance methods seem capable of addressing black swans. Like stated by Aven & Krohn (2014): "we need a broader concept of risk to make risk management meaningful in a black swan world". This means that even more than for other risks, dealing with the tolerance of black swan risks will become an issue of finding a suitable and a very holistic decision-making process where risk management solutions based on increased resilience can be taken into account as a cost-effective way to address a number of black swan risks simultaneously.

4 Synthesis and conclusions

The main conclusion concerning risk acceptance is that basing risk acceptance only on a single quantifiable measure does not sufficiently take into account the multiple dimensions of risk. Therefore, while the various indicators presented in the previous section - such as VSL and WTP - may serve as inspiration, the final decision on risk acceptance should be made in a holistic manner in line with the modern risk perspective, paying attention on things such as the assumptions taken and the strength of knowledge. A holistic and robust decision-making process becomes the main focus, and this includes what Aven calls the managerial review.

Another related conclusion is that it is probably better to try to look at various risks in relation to each other and to judge their acceptability through this holistic view, rather than comparing individual risks with a fixed acceptance limit. This is supported by the observation that different sectors will require different acceptance criteria and these criteria in society are not derived from mathematical formulas but are rather the results of a complex social process. Visualizing all relevant risks together in the same risk picture become thus one key objectives. Ideally, the cost (at large; not only financially) of the various risk treatment options could also be visualized together with the risks so that the various options could be compared with each other and related to the expected positive impact in terms of risk reduction. The risk picture could then be used to produce answers to both questions of Paté-Cornell: which risks are deemed acceptable, and what are the measures that are the most justifiable in any given situation. It is acknowledged that some risks could be judged simply inacceptable mainly on ethical grounds irrespective of their position compared to other risks.

Concerning risk aversion, there may be different preferred policies in different agencies/companies/organizations and at different times, so it would be unfortunate if the risk aversion policy needed to be locked once for all as a part of the risk management framework. Ideally, different policies can be implemented, and the impact of different risk aversion strategies could be simulated.

Besides leaning on proper methods and rational reasoning, the decision-making process will have to consider public opinions, political interests and other non-quantifiable aspects. The process will undeniably be exposed both to the human biases and limitations of the decision makers and the – possibly at least partly irrational - demands of the public. Hopefully, awareness of these vulnerabilities and adapted working methods can limit the negative impact.

How black swans and very-low-probability risks enter into the acceptance process is a special challenge which needs to be addressed, and obviously, their acceptance must not be based on probability alone.

In terms of the ethical basis for a risk management framework, one can imagine that different approaches could be taken. Therefore, the more flexible the framework is in adopting different ethical priorities, the better. From the point of view of a national transport safety agency, it seems obvious that both individual and societal risk perspectives need to be maintained. It would be difficult to justify imposing much higher risks on some groups of individuals than others. On the other hand, risk distribution in the society will realistically always be somewhat uneven. Once again, safety and risks are not the only criteria in societal decision making.

The same preference for flexibility applies to the various industry requirements and priorities, such as the EASA priorities mentioned in the previous section: it is an advantage if risks can be viewed flexibly with or without such weighting factors to compare potential differences from the point of view of risk acceptance/prioritization.

This chapter completes the thematic review of literature. In Chapter 6 the focus moves to the level of the risk management framework and process, combining various topics reviewed so far.

Chapter 6 – State of the art in transport risk management frameworks

The objective of this chapter is to review existing risk management frameworks and related methods and guidelines. The aim is on one hand to review the state-of-the-art, and on the other hand to study how aspects of existing methods and guidelines could contribute to the NRMF. Another major task in this chapter is to consolidate all requirements for the NRMF from Chapters 2-5 and to structure them logically.

Section one reviews guidelines for risk management frameworks and their components which have not yet been discussed in the previous chapters. The structured list of requirements for the NRMF is developed in section two.

Section three focuses on current risk management methods within the industry. It is deemed important to review these methods because the developed risk management framework should eventually be able to replace some of the existing methods. As already mentioned in Chapter 2 and Chapter 4.1.3, the industry methods have not evolved a lot in the recent past. They tend to cover the whole process from the beginning to the end because they are rooted in the pragmatic needs of the industry. However, they may have a weak scientific backbone or ignore some of the major challenges. The review is started with the methods from the industry, because this makes a better chronology in terms of how well the methods embrace modern scientific concepts.

Section four covers current risk management frameworks in scientific literature. These methods tend to be more up-to-date with aspects such as the new risk perspectives. However, they often cover only parts of the full risk management process.

1 Guidance for the risk management process

This section focuses on material in the literature at the level of the risk management framework and/or process. The discussion starts with general advice for the risk management process and then addresses decision-making, interventions with the transport system and finally continuous learning. The aim is to capture any significant contributions which have not been captured so far because the focus has not been at the level of the process.

Paté-Cornell (2002) proposes that the goal of risk management could be "to support the optimal use of limited resources to provide the best protection to the maximum number of people". On the other hand, there are other priorities than safety and cost in an organization, so risk management should address all these priorities. Kontogiannis et al. (in press) discuss the principles, processes and methods for *total safety management*. They argue that risk management should be part of all decision-making and organizational processes and *create value for the business*. A key element is a *common operational picture about risks*. Performance monitoring, operational feedback and participation of all stakeholders are seen as key success factors. The same idea *to integrate risk assessment with other key organizational processes* is also promoted by Thekdi & Aven (2016). They state that there is a need for a *risk-performance framework* in line with the new risk perspective, addressing uncertainty as a main component of risk, as well as the knowledge dimension and surprises.

Aven & Krohn (2014) dedicate the appendix A for presenting the key features of the new thinking about risk. The part A.2 covers risk assessment and risk management with the following points:

1. Risk management covers all activities implemented to manage risk and is concerned with balancing value generation and avoiding the occurrence of undesirable events.
2. A risk assessment describes risk for various alternatives, identifies key risk contributors and factors and compares the results with relevant reference values. A risk assessment supports decision making on where to reduce risk and what alternative to choose.
3. The results of a risk assessment need to be put in a wider decision-making context, which we can call a managerial review and judgment process. This process takes into account the limitations of the assessment and incorporates other concerns not addressed in the assessment.
4. The cautionary and precautionary principles have an important role to play in risk management, to ensure that the proper weight is given to uncertainties in the decision making.
5. Robustness and resilience are examples of cautionary thinking.
6. Risk acceptance should not be based on the judgements on probability alone.
7. Probability-based risk acceptance criteria should not be used.
8. Risk reduction processes are recommended based on the ALARP, using ideas presented in Aven (2013a), which give due attention to uncertainties and the strength of knowledge supporting the probabilistic analysis.
9. Cost-benefit type analyses need to be supported by risk assessments to provide adequate decision support, as these analyses are expected value-based which, to a large extent, ignore risks and uncertainties.

Pate-Cornell & Cox (2014) propose some foundations for better risk management. These points include:

- Assess the urgency and value of information. Is collecting or waiting for additional information before acting actually more costly than it is worth?
- Anticipate, monitor and prepare for rare and not-so-rare events. It is proposed that the best strategy for black swans may be to monitor for singles of unusual occurrences and to put in place a resilient structure of organizational connections, financial reserves, and access to human intelligence and knowledge that allows for quick, creative local responses.
- Test and learn deliberately. Test key assumptions, try to learn from experience and to capture data and lessons for future reference.

Aven (2008) presents the idea of *semi-quantitative approach to risk analysis*. Some aspects of the risks are thus quantified and some are not. The key points of this approach are:

- Establishing a *qualitative risk picture*, containing the hazards, threats and accident scenarios, as well as some other factors such as barriers, possible risk reducing measures, uncertainties, vulnerabilities and manageability factors.
- Using the risk picture for *rough risk categorization*, reflecting probabilities/frequencies of hazards/threats, expected severity of the above if they occur, and uncertainty factors.
- Based on the above, making judgments about risk acceptance and *comparing alternatives*.

Villa et al. (2016) recognize the challenge of risk management in an environment which is constantly evolving. They promote the development of novel approaches to risk assessment and management which would better *consider the dynamic evolution of conditions, both internal and external to the system, affecting risk assessment*.

Duijm (2015) discusses the challenges and problems in the use and design of risk matrices. His paper has the great merit of drilling into the very practical consequences of the conceptual challenges behind risk assessment, and behind the use of risk matrices in particular. The paper highlights the very same challenges that the author of the current dissertation faced during the design phase of the risk matrices for the developed RMF, in 2013-2015. Duijm mentions several typical problems related to risk matrices such as the clarity of the textual descriptions attached to different categories, the limited resolution in matrices, compatibility with other matrices (e.g. used at the corporate level) and the various biases like the so-called "centering bias". He highlights several noteworthy considerations and recommendations:

- Both the consequence scale and the probability scale must cover the full range relevant to the assessments. The use of logarithmic scaling for both consequence and likelihood scales is

recommended in order to cover the several orders of magnitude typically required. It is also recommended to use the same scaling for both dimensions.

- Due to different possible interpretations of category definitions, it is advisable that nominal categories are linked to some quantifiable reference, e.g. numerical ranges.
- Instead of using discrete categories it is also possible to produce a *continuous probability consequence diagram*. This approach has several advantages, e.g. the resolution problem related to large categories disappears.
- It is desirable to be able to *present the level of uncertainty*. If a continuous probability consequence diagram is used, uncertainty can be symbolized by using boxes or bands instead of single points, thereby showing that consequence and probability values are estimated ranges rather than precise values.
- Because risk matrices need to be customized to the specific circumstances, it is not advisable to use a single standard risk matrix over different activities - e.g. at the corporate level. What is tolerable at company level might not be tolerable at department level, or vice versa.
- An important detail is *what is used as the reference consequence* in the risk assessment, and does the estimated probability refer to that same consequence. This same problem was described in the ARMS (2010) document with the two questions: "severity of what? Probability of what?" Duijm presents three different alternatives for this. Roughly speaking, the reference is either the worst-case consequence, the most likely/representative damage, or different outcomes are considered separately with different respective probability values.
- Event severity typically has several dimensions, and therefore the different types of impact cannot be *a priori* directly compared or presented on the same scale. This would also make the use of a single matrix impossible.
- *Aggregation of risks* is a related issue. How to combine the impact of a single event on different areas of concern, e.g. financial, environmental and safety? Or, how to combine the risk of multiple hazards that originate from a single activity? As a general rule, it is advised that the need for aggregation should be avoided. It is also stated that low risks cannot be cumulated together and considered equivalent to a single (event with) high-risk.
- Overall, Duijm concludes that due to the several serious disadvantages of risk matrices, *quantified risk assessment should be preferred*, e.g. the continuous probability consequence diagram.

Concerning *decision-making* within the risk management process, Paté-Cornell (2002) proposes the following elements as the bases for a good decision process:

- "A sound *legal basis* with clear understanding of individual and societal risks, burden of proof, and treatment of economic effects (including when and how cost and benefit considerations are legally acceptable).
- A *monitoring system* that allows early detection of chronic problems, hot spots, repeated accidents, clear threats, etc.
- An *information system* including a risk analysis with appropriate characterization and communication of uncertainties and assumptions.
- A *communication system* such that this information can circulate among and be fully understood by concerned individuals and organizations.
- A sound *criterion for selection of experts* and a *mechanism of aggregation* of expert opinions that reflects the characteristics of the problem.
- A *public review process* in which the information used and the risk analysis method can be examined and criticized by members of the public, industry, interest groups, etc.
- A clear but flexible *set of decision criteria* that reflect public preferences given the nature of the hazard, the state of the information, and the economic implications of considered regulations.
- An appropriate *conflict resolution mechanism* (mediation, arbitration, etc.).
- A *feedback mechanism* such that data are gathered post facto and used in an appropriate and predictable way to measure the regulatory effects, including those that may have escaped initial policy analysis."

It can be observed that most of these elements are also explicitly part of the ISO 31000 risk management framework.

Aven et al. (2007) present a decision framework for risk management. The decision process consists of four steps:

1. Framing: describe objectives and define the problem.
2. Alternatives: generate alternatives, select the method, evaluate alternatives.
3. Managerial review and decision
4. Implementation of decision: implement, evaluate.

Due to the difficult value judgments in decision-making, involving uncertainty, Aven et al. promote the view that *decisions should be taken by people* during the managerial review and it is thus not desirable to develop tools that prescribe or dictate the decision - such an approach would be too mechanical. Johansen & Rausand (2014) express the same view by stating that decision-making should be *risk-informed* rather than risk-based. In terms of risk acceptance, Aven et al. (2007) point out that to be precise, in accepting a risk, *what is really accepted is a solution with all its attributes*. The values and goals of the stakeholders are integrated into the decision-making, in line with the process presented by Aven & Vinnem (2005) and Ersdal & Aven (2008). Aven & Vinnem (2005) make the following observations related to the decision-making process:

- A key success factor is to have a drive for *generating alternatives*. This requires a good climate for considering possible changes and improvements.
- A critical step is for the analysis team to be able to *communicate the information surrounding the analyses to the decision-makers*. The decision-makers must fully understand what the analyses and the analysts express. A direct communication without any filters is recommended.
- A clear *visualization of the decision-makers trade-offs* between safety and other aspects could be useful.

Moving on to the interventions with the transport system, the first topic to review is adaptive policy making. Swanson et al. (2010) start their reasoning from the observation that policies crafted to operate within a certain range of conditions often face unexpected challenges outside of that range, leading to unintended impacts and failing to accomplish the original goals. They promote *adaptive policies* which are designed to function more effectively in complex, dynamic and uncertain conditions, and are therefore particularly well suited for complex adaptive systems. The idea is to develop policies that are not targeted to be optimal for a best estimate future, but robust across a range of futures. The inevitable policy changes become part of a larger recognized process instead of being forced to be made repeatedly on an ad hoc basis. The decision-making process combines learning and action. Swanson et al. (2010) present a collection of intervention principles for complex adaptive systems:

- Policy set-up:
    - Respect history
    - Understand local conditions, strengths and assets
    - Understand interactions with the natural, built and social environment
    - Look for linkages in unusual places
    - Determine significant connections rather than measure everything
    - Public discourse and open deliberation are important elements of social learning and policy adaptation
    - Build trust, collaboration, consensus, identity, values, hope and capacity for social action
    - Use epistemic communities to inform policy design and implementation
- Policy design and implementation:
    - Create opportunity for self-organization and build networks of reciprocal interaction
    - Ensure that social capital remains intact
    - Promote effective neighborhoods of adaptive cooperation
    - Members of the population have to be free and able to interact
    - Facilitate copying of successes

  - Clear identification of the appropriate spatial and temporal scale is vital to integrated management (the ecosystem approach)
  - Match scales of ecosystems and governance and build cross-scale governance mechanisms.
  - Promote variation and redundancy
  - Encourage variation
  - Balance exploitation of existing ideas and strategies and exploration of new ideas
- Monitoring and continuous learning and improvement:
  - Integral to design are the monitoring and remedial mechanisms — should not be ad hoc additions after implementation
  - Fine-tune the process
  - Learning and adaptation of the policy be made explicit at the outset and the inevitable policy changes become part of a larger, recognized process and are not forced to be made repeatedly on an ad hoc basis
  - Policies should test clearly formulated hypotheses about the behavior of an ecosystem being changed by human use
  - Learn to live with change and uncertainty
  - Policies should be expected to evolve in their implementation
  - Increase information on unknown or partially unknown social, economic and environmental effects
  - Evaluate performance of potential solutions and select the best candidates for further support
  - Understand carefully the attribution of credit

Swanson et al. (2010) go on to present "seven tools for the adaptive policymaker". The aim is to be able to create policies which can adapt to both anticipated and unanticipated conditions. These seven tools (which are in fact rather methods or approaches) are summarized below:

1. *Integrated and forward-looking analysis*. After crafting the policy, key factors driving the policy performance will be identified. The policy can then be tested in a "policy wind tunnel" using defined scenarios and estimating the impact on the performance. If weaknesses are discovered there are at least three options for improvement:
   a. developing a policy option that performs in a range of anticipated future conditions
   b. Mitigating or hedging against the adverse or unintended impacts that can be foreseen
   c. Identifying how the policy might need to be adjusted in the future, and defining how such adjustments will be triggered.
2. *Built-in policy adjustment*. When the need for future policy adjustment can be anticipated, fully- or semi-automatic adjustment mechanisms can be built in the policy.
3. *Formal review and continuous learning.* It is recognized that it will be necessary to refine the interventions from time to time, typically based on stakeholder feedback. A specific way to implement this principle is the use of so-called policy pilots, i.e. a phased introduction of policies or programs, allowing them to be tested, evaluated, and adjusted when necessary, before being rolled out in large scale.
4. *Multi-stakeholder deliberation.* Based on both literature and practical experience, the following guidelines are offered:
   a. Participation is voluntary
   b. The effort is structured and led by skilled facilitators
   c. The process is guided by explicit rules and procedures
   d. All participants have an opportunity to speak, and all should feel that their views have been heard and considered without risk or prejudice
   e. Participants include a broad range of stakeholders directly or indirectly affected by the decision
   f. Deliberative proceedings are transparent and accessible.
   g. Participants engage each other on the basis of communication and open discussion.

   h. Deliberation is aimed at an explicit decision context. It is not intended merely to generate opinions.
   i. Deliberation is most effective when conducted face-to-face.
5. *Enabling self-organization and social networking*. The value of a strong social network shows on one hand in the resilience that well-linked communities have, and on the other hand in the capacity to create solutions to problems spontaneously without any external input or formally organized interventions. Policies can promote self-organization and networking by trying to remove any resource barriers, creating effective spaces and issues for adaptive cooperation, ensuring that social capital remains intact, and facilitating copying of good practices.
6. *Decentralization of decision-making.* Importantly, when decision-makers are closer to the people affected, they get faster and better (formal and informal) feedback and can implement policies more flexibly.
7. *Promoting variation.* Implementing a variety of policies to address the same issue increases the likelihood of achieving desired outcomes. Therefore, introducing small-scale interventions for the same problem offers greater hope of finding effective solutions. Due to the nature of complex systems, many interventions inevitably fail – and the failures need to be seen as means for finding successful interventions. The variation can be introduced through: design (e.g. through several parallel experiments), facilitating an environment for variation to occur, and/or through using feedback to create variation.

Swanson et al. (2010) point out that *policies that fail* will not only fail to achieve their desired objectives, but *often actually make things worse*. The final aims for adaptive policy making are policies which adapt to anticipated and unanticipated conditions, but also enhance local resilience to unforeseen events in general through stakeholder participation and commitment. Working on adaptive policies, especially for the feedback and refinement, allows people to gain experience in a variety of policy approaches and their success in different conditions.

Marchau et al. (2010) discuss dynamic adaptive policies in the context of transport policy making. They argue that traditional scientific approaches have serious shortcomings in handling deep uncertainty regarding long-term transport policy making in an appropriate way. They point out that one way to exercise adaptive policy making is to start the implementation phase prior to the resolution of all major uncertainties and to adapt the policy gradually over time based on new knowledge.

Adaptive management and more specifically adaptive risk management breaches the gap between risk assessment and adaptive policies (and other interventions). Adaptive management is based on the notion that in some conditions no single best policy can be selected but rather a set of alternatives should be dynamically tracked to gain information about the effects of different causes of action (Linkov et al. 2006). Bjerga & Aven (2015) illustrate how the principles of adaptive management can be applied in risk management, leading to *adaptive risk management*.

Grote (2015) offers an interesting and a somewhat unusual perspective by proposing that for improved risk-related decision making, *the level of uncertainty might have to be deliberately increased*. Typical engineering approaches would aim at reducing uncertainty, and eventually trying to maximize stability and centralized control. However, first there are limits to being able to reduce the uncertainty due to the nature of complex systems. Secondly, to use the full potential within a complex adaptive system (and to increase its resilience), there should be room for self-organization and innovation, but this requires decentralization and thus increases uncertainty. Grote promotes making deliberate choices as to the level of uncertainty in order to strike a balance between stability and flexibility in the organization, and also to match the level of control and accountability that people have. A very practical example for increasing flexibility (and uncertainty) is to phrase rules and procedures with some options to choose from or with parts that may or may not be applicable based on the conditions. Finally, she joins the many other authors already mentioned who promote the importance of humans taking the key decisions by stating: "Empirical evidence has been accumulated to show that the prerequisites for mathematical models of rational choice are often not met in actual decision-making". Perrow (1984) reminds about another

weakness of risk assessments: "Risks from risky technologies are not borne equally by the different social classes; risk assessments ignore the social class distribution of risk".

It is clear that when the target system is complex, the methods to be used and the solutions need to be adapted and would often not be simple and straightforward. This introduces the challenge of communicating the problems and solutions to different stakeholders. Moxnes (2000) points out that complexity is among the top five attributes explaining diffusion, or in the case of complexity rather the lack of it. For someone to evaluate the appropriateness of a solution, one needs to understand why it is effective, and if the explanation is too complex, the explanation will be just as confusing as the reality itself. The challenge is even bigger if the audience is large and diverse or consists of people with little detailed knowledge and weak incentives for learning. When institutions are in place and actors have appropriate incentives to make good decisions, the situation is better.

The presented analysis and solutions should be in line with the mental models and heuristics of the audience, or there needs to be an attempt to change their mental models and heuristics. This relates to what Sterman (1994) calls double-loop learning: the point is not only to be able to adjust inputs to the system hoping to obtain the wanted response, but also to gain a better understanding of the true dynamics of the system and update one's own mental models accordingly (which should then enhance the success of future interventions with the system). Based on the discussion on learning in the context of complex systems, in Chapter 3.5.3, it should be clear that it is very difficult to replace people's simple mental models with very complex models relating to complex adaptive systems.

Finally, in the context of adaptive risk management and learning, Cox (2012) offers some questions which may stimulate finding new and better solutions for risk management and decision-making:

- Instead of (or in addition to) asking "What can go wrong?" one might ask "Is there a clearly better risk management policy than the one I am now using?"
- Instead of asking "How likely is to happen?" one can ask "How probable should I make each of my next possible actions?"
- Instead of asking "If it does happen, what are the consequences likely to be?" one can ask "Would a different choice of policy give me lower regret (or higher Expected Utility of consequences), given my uncertainties?"
- Instead of asking "What should I do?" one might adopt a multi-agent perspective and ask "What should we do, how might they respond, and how should we respond to their responses?"

2 Refined objectives for the framework

Based on the whole literature review, it is now possible to establish the detailed requirements for the NRMF and the principles that it must reflect. These are first collected thematically from each chapter, and then consolidated and re-arranged in order to establish a well-structured set of requirements.

Embracing the modern risk perspectives means that:

- *Uncertainty* and *strength of knowledge* (SoK) are in central roles. There needs to be a specific focus on *building knowledge* prior to risk assessments and decision-making.
- *Assumptions* and the (assumed) *strength of knowledge* are integral parts of the results
- The *black swans* need to be addressed and the high-impact–low-probability scenarios must not be dismissed solely due to their low probability.

Treating the transport system as a complex adaptive system (CAS) has the following consequences:

- The transport system is seen as *a system of sociotechnical systems*, with all the typical features of CAS discussed in Chapter 3, including non-linearity, unpredictability and counter-intuitiveness.

- In terms of trying to understand the system, besides the usual *analytical* perspective, the transport system (and its subsystems) should also be seen as parts of a larger system - adding a perspective of *synthesis*. System diagrams may also be useful.
- *Humans* are invaluable in trying to make sense of complex systems. Computers and rigid information management schemes cannot cope with some important aspects such as multiple diverse perspectives, value judgments, understanding of context and overall objectives.
- Because information and knowledge are distributed all over the system, and in order to create diversity, *people from outside the core group* (in the agency) and *operational people* from the field should be pulled into the process. Besides the core group, *variation* in the group composition should be created.
- When preparing interventions, it should be taken into account that different risks are interconnected (Ackoff's "*mess*"). A holistic view of the whole must take precedence over fragmented approaches. The RMF should allow treating several issues together as an entity whenever possible, e.g. a solution to one safety issue may degrade the situation for another.
- When trying to influence the system, one must accept that it is *not directly controllable* by anyone and that *unintended consequences are typical*. Therefore, feedforward prescriptions of action are risky. *Adaptive approaches* such as experiments (with feedback and adaptation) are recommendable.
- The ways to interact and the attempts to influence the system need to be *adapted to the nature of the system* (or sub-system). The Cynefin-framework can offer a helpful reference.
- A complex system can never be fully understood, but the risk management process should facilitate constant learning *about the transport system, its risks and its reactions to different types of interventions*. Such *coevolution* can be seen as one of the key objectives of the RMF and reflect positively in the capability to influence the system positively.

The conclusions related to safety imply that:

- The Safety-II paradigm is embraced and there is an attempt to support the concept of safety as something positive – as the presence of something – and *measure safety on a positive scale*. One consequence is the desirability to try to understand the *level of resilience* at different organizations, e.g. operators.
- As the transport system is constantly evolving, and so are the risks within, the RMF needs to be supported by a continual flow of data/information from the system, allowing continuous learning about the system at the level of events, organizations and the whole system.
- There is already an established flow of *safety information* within the transport system. The RMF needs to be able to digest the types of safety data used currently – including individual safety events - as an input and transform such data into useful risk information. This should improve the total view on risks and support decision making.
- It should be possible to establish Safety Cases, i.e. risk assessments on specific issues and (planned) changes, and obtain meaningful results.
- Safety and risk management functions need to be carried out within acceptable levels of *efficiency, resource and cost*.
- Humans are subject to many *limitations and biases*, affecting also the capability to assess risks and take related decisions. These human properties cannot be taught away but there should be an effort in the design of the RMF to try to counteract these effects. Possibilities include diversity of people involved, involvement of "outsiders", encouraging dissenting views, use of adapted tools and models, embracing uncertainty openly, etc. The risk management process should facilitate such practices systematically.

Combining the Safety-II approach with the CAS-perspective supports the following points:

- *Safety is as an emergent property in the system*. This means that it cannot be created simply by executing a recipe. This also means – due to the laws of CAS - that it cannot be managed directly.

- Safety cannot be considered apart from *other priorities in the system*. There are links, influences and trade-offs with other priorities, such as environmental sustainability and economic considerations. Therefore, the RMF should offer opportunities to *balance the different priorities* in the context of risk management.
- The previous point suggests that it should be possible to *present different types of risks* (financial, reputation, etc.) through the same RMF which is used for safety risks.
- The RMF has to consider and act at *different levels*: events (system behavior), organizations (elements) and a system level (be it the whole system or a part of it).

Review of the risk acceptance topic gives the following requirements:

- Final decision on risk acceptance should be done in a *holistic manner*, including the aspects of the modern risk perspectives, rather than compared to a single numerical risk acceptance criterion. This is also supported by the fact that different modes of transport have different levels of acceptable risk, and even within one mode there may be different limits, especially when there is both commercial and private activity.
- Because risk acceptance becomes multi-dimensional, the whole *decision process* needs to be carefully designed.
- The RMF should facilitate presenting all risks/threats in a single risk picture constructed in such a way that it helps compare risks *relative* to each other. In order to allow true prioritization, the risk picture should be able to host the risks of all modes of transport.
- *Risk treatment options* should also be presentable in the risk picture with some measure of their associated "costs" (in a large sense), so that *alternatives can be compared* (e.g. which treatment options of which risks would seem to bring the best overall result).
- Ideally, the risk *aversion policy should be left flexible* (i.e. not pre-determined).
- There needs to be a way to *present and properly address black-swan-risks* and other very low probability risks, not based solely on probabilities.
- The RMF should be flexible in being able to host different ethical principles. The framework should reflect solid ethical values.
- Due to the need to integrate different kinds of aspects and priorities (e.g. in the managerial review), as well as due to the need to refer to ethical principles, the importance of *human judgment* comes up again, as opposed to some pre-defined criteria.
- If there are specific industry references related to risk acceptance (or severity assessment), it should be possible to reflect those in the RMF.

Review of existing risk management frameworks and related guidelines highlights the following points:

- *Risk management needs to be integrated with the overall decision-making process* and take into account all the *different priorities, not only safety*.
- Creating a *holistic risk picture* is recommendable
- Decisions and judgments cannot be left to algorithms but need to be *made by humans* who can try to integrate together all the different priorities and value judgments.
- *Stakeholders* need to be involved to give their inputs and feedback.
- An *adaptive approach* needs to be adopted for risk management and interventions. This means generating a lot of different alternatives and testing them, preferably in small-scale, before gradual larger implementation.

In addition to the above requirements derived from the literature research in Part I, there are a few more requirements that are related to the context of a national transport safety agency. As described in more detail in Part III, the author cooperated closely with the Finnish transport safety agency during two years for the development of the RMF. The following requirements emerged during that time:

- There are *proposals for action and/or proposed priorities coming from outside the risk assessment process*, typically coming from the political system, e.g. from the ministry of transport. Some actions are also imposed by international (or European) organizations like ICAO, IMO or EASA. There must therefore be a way to integrate such items (with their

associated risks) into the risk picture and into the risk management process and to see them in relation to other risks and proposed actions.

- In terms of interventions with the transport system, there are several agencies with specific limits of authority, and interventions with the transport system might often require acting across these inter-agency borders. Therefore, *interventions should not be limited to actions which are within the powers of the agency*, but rather when multiagency intervention is required, one of the intervention modes should be to *convince* other agencies to act in concert and to *coordinate* such interventions. The same applies to interventions where cooperation is needed with other types of organizations, such as commercial operators.
- People within the agency possess a lot of relevant knowledge. The RMF and the associated working process should enable capturing this knowledge, some of which is likely tacit knowledge.
- If the agency is sensitive to several types of priorities (e.g. safety, environment, transport businesses' interests), then the severity of the potential outcomes of safety events will also be multi-dimensional: there may be for example injuries to people *and* environmental damage. If these different dimensions are accounted for on separate scales, the comparison of total severities will be very difficult. It would therefore be preferable to be able to measure *different types of severities on a single scale*: the total severity value would consist of several components (human losses, environmental damage, material damage, etc.).

Very importantly, the workload pressure on existing resources was very tangible and this put a lot of weight on the (already listed) requirement to be able to carry out the process with as few resources as possible. More specifically:

- The method needs to help the agency focus its resources as precisely as possible on the points in the transport system which need intervention most, based on risk assessment. For example, which *organizations* and which *threats*?
- The low-resource (and lead time) requirement becomes even more critical for steps which need to be repeated very frequently, e.g. typically any processing related to incoming raw safety data, due to its high volume.

This last requirement creates a tension with the earlier requirement that all such data should be fed as inputs to the risk management process and transformed into risk information.

Some of the above points are very tangible requirements while some may seem abstract. However, they all have practical implications on the RMF. Some of these requirements can be merged together to create a compact and logical set of requirements. They can also be reorganized based on the ISO 31000 process stages. The result is the following set of requirements:

Risk Identification

- (RI-1) As the transport system is constantly evolving, and so are the risks within, the RMF needs to include provisions to feed in a continual flow of data/information from the system.
- (RI-2) There is already an established flow of *safety information* within the transport system. The RMF needs to be able to digest *the types of safety information* used currently and transform such data into useful risk information.
- (RI-3) The need to carry out a risk assessment related to a safety issue or a change must be one of the normal inputs to the process.

Risk Analysis

- (RA-1) The new risk perspective (incl. focus on uncertainty and surprises) should be embraced.
- (RA-2) There needs to be a specific focus on *building knowledge* prior to risk evaluations and decision-making.
- (RA-3) Knowledge of the *people within the organization* (e.g. the safety agency) needs to be captured, including tacit knowledge.

- (RA-4) *People from outside the core group* (in the agency) and *operational people* from the field should be pulled into the process (to capture knowledge and to create diversity).
- (RA-5) One should aim at understanding the *resilience* of the organizations carrying the risks, and to find ways to describe it, preferably in a comparable way.
- (RA-6) The RMF should feature a *holistic risk picture* including risks/threats from the 4 modes of transport and allow comparisons of risks *relative* to each other.

Risk Evaluation

- (RE-1) The decision process should be such, that the decision on the acceptability/priority of a risk or a solution can be done in a *holistic manner*, *taking into account the assumptions* and the *strength of knowledge*, while paying attention both on the risk and on the *costs of its treatment* and considering the *various priorities, not only safety*.
- (RE-2) Consequently, ideally the risk picture should be able to host non-safety risks (e.g. financial, reputational risks) in a compatible manner.
- (RE-3) Decisions on risk acceptability need to be done by humans, due to their capability to consider ethical principles, various conflicting priorities and make value judgments; i.e. the risk management methodology does not need to (and should not) try to produce the final answers directly.
- (RE-4) The *black swans* need to be addressed and the high-impact–low-probability scenarios must not be dismissed solely due to their low probability.
- (RE-5) Ideally, the risk *aversion policy should be left flexible* (i.e. not pre-determined).
- (RE-6) It should be possible to apply specific industry references (e.g. such as by IMO, RSSB or EASA) related to risk acceptance or severity assessment (e.g. "1 fatality corresponds to 10 severe injuries").

Risk Treatment

- (RT-1) *Risk treatment options* should also be presentable in the risk picture with some measure of their associated "costs" (in a large sense), so that *alternatives can be compared*.
- (RT-2) There needs to be a way to *present and properly address black-swan-risks* and other very low probability risks, e.g. by building suitable *resilience* in the system.
- (RT-3) The interventions with the system need to be *adapted to the nature of the system* (or sub-system), e.g. referring to Cynefin.
- (RT-4) As the transport system is seen as *a system of sociotechnical systems*, it is important that interventions embrace the typical features of CAS, including non-linearity, unpredictability, counter-intuitiveness, unintended consequences of interventions and the fact that emergent system properties like safety cannot be controlled directly.
- (RT-5) An *adaptive approach* needs to be adopted for risk management and interventions. A recommended *modus operandi* is to use parallel experiments (e.g. small-scale implementations) and to adapt them based on the feedback e.g. by expanding the successful interventions.
- (RT-6) A holistic view of the whole must take precedence over fragmented approaches because in a CAS, different risks (and interventions) are typically interconnected (Ackoff's "*mess*").
- (RT-7) *Stakeholders* need to be involved systematically to give their inputs and feedback.
- (RT-8) While a complex system can never be fully understood/described, constant learning *about the transport system, its risks and its reactions to different types of interventions* should be facilitated. Such *coevolution* between the decision makers and the system can be seen as one of the key objectives of the process and improve future interventions.
- (RT-9) Irreplaceable human characteristics (e.g. sensitivity to history, context, objectives, values) should be exploited in making sense of the CAS, but distinct features in the process should counteract human biases and limitations (e.g. availability heuristic, groupthink).
- (RT-10) Both from the intervention and learning points of view, three distinct levels need to be considered: events (system *behavior*), organizations (elements) and the system level (parts of, or the whole transport system); and besides *analysis*, there should be an effort to apply *synthesis*,

i.e. study the role the system plays within the larger context, and the objectives/constraints from above.

Context

- (C-1) Safety and risk management functions need to be carried out within acceptable levels of *efficiency, resource and cost*. This is particularly critical for process steps which are performed frequently (e.g. event risk assessment).
- (C-2) Thanks to the process, the Agency should be able to focus its resources *as precisely as possible on the most critical risks* (which could be combinations of risk scenarios and organizations, for example).
- (C-3) In addition to inputs from Risk Identification, it should be possible to integrate into the process *proposals coming from outside*, e.g. from the political system, and make sure such items are comparable with other elements.

The development and validation of the RMF is based on this set of requirements.

3 Risk assessment practices in the industry

The review of current risk management practices in the industry is started with methods used at the operational level, typically promoted and/or imposed by international roof organizations. Some of the below methods have already been referred to in the context of current risk concepts (Chapter 2) and in the context of safety management systems (Chapter 4.1.3).

The *Safety Management Manual (SMM) of the International Civil Aviation Organization* (ICAO 2013) covers safety risk management, both for operators and state safety agencies.

- Steps covered: identification of hazards and their potential consequences, probability and severity assessment, assessment of risk level and tolerability, decision to take action and to continue operation or cancel the operation.
- Inputs: data and information on hazards
- Outputs: risk levels, decisions on risk tolerability and mitigation.
- Relationship with new risk perspectives: none.
- Relationship with complexity: none.
- Risk acceptance approach: three levels: acceptable, unacceptable and in between “acceptable on risk mitigation”. Based directly on the risk level from a matrix.
- Principal elements of the method: generic flowchart of the process and a few matrices with brief descriptions on the overall process.
- Description of the method:

  The provided guidance for risk management is superficial, remaining at the level of headlines. One flowchart describes the overall flow (from hazard identification to decisions) and includes only a few keywords. Two 5-level tables are offered to describe different severity levels in broad terms, as well as a similar 5-level likelihood table, 5x5 risk table and a tolerability table. There is no advice neither at the level of risk concepts (e.g. aversion, uncertainty) nor at the level of practical solutions (e.g. how to transform hazard data into risk information) – i.e. all real questions about risk assessment and management are avoided. The guide does acknowledge the need to feed hazard data into the risk management process. Other ICAO material (ICAO 2009, pp. 51-61) illustrates that there is fundamental confusion about risk analysis (e.g. applying it to an accident which already took place). The Reason model still acts as the principal conceptual framework for safety management in the SMM.

  The requirement given to state safety agencies is to “ensure that service providers implement the necessary hazard identification processes and risk management controls”. This includes

defining "a mechanism to agree with individual service providers on acceptable safety performance levels to be achieved through their SMS" (Chapter 4.2.16, p. 71).

The Safety Management International Collaboration Group (within aviation) published the *Risk based decision making principles* (SMIGC 2013). The document does not deliver a real risk management framework but proposes high level principles for carrying out the activity.

- Steps covered: identification of hazards and their potential outcomes, probability and severity assessment, risk mitigation strategies.
- Inputs: safety data
- Outputs: potential risk mitigation strategies
- Relationship with new risk perspectives: none
- Relationship with complexity: none
- Risk acceptance approach: data-driven, probability-severity based.
- Principal elements of the method: calculating risk value, recommend urgency of risk control, evaluate risk control options.
- Description of the method:

  Most of the document (pages 3-13) is dedicated to safety data, data management and hazard identification, which once again underlines the high importance given to such data in aviation safety work. Risk management is covered in pages 14-16. Risk is defined as the product of severity and probability. Use of quantitative data in risk assessment is preferred as "it tends to be more objective". Advantage of team work in risk analysis is recognized and the background of the experts is seen as a key driver for the quality of the analysis. Statistical or observational data is considered important and if judgmental inputs are used, "they should be expressed in quantitative terms". Risks are either tolerable or not, and in the latter case potential mitigation strategies need to be evaluated. Different "risk mitigation approaches" are introduced, and various selection criteria are mentioned.

The International Maritime Organization (IMO) leans on the *Formal Safety Assessment (FSA) methodology*. As its name suggests, the FSA is more a one-time assessment of risks than a tool for continuous risk assessment & management in an operational setting. FSA is reviewed here because unlike the ICAO SMM content, it does specify risk analysis standards in detail, and thus reveals the underlying thinking in the maritime world.

- Steps covered: identification of hazards, risk analysis, risk control options, cost-benefit assessment, recommendations for decision-making.
- Inputs: problem definition and a generic model describing the relevant functions and attributes related to the problem in question.
- Outputs: risk control options, cost benefit assessments and recommendations, in the form of a report.
- Relationship with new risk perspectives: risk is defined as the combination of the frequency and severity of the consequence, so the approach is classic. However, uncertainty is recognized in the form of "confidence" per different areas of the risk model.
- Relationship with complexity: there is a conscious effort to address the full organizational scope relevant to the study. However, the system is not treated as a CAS.
- Risk acceptance approach: FN-curves, ALARP
- Principal elements of the method: use of fault and event trees to build a risk model, FMEA, HAZOP, risk index, FN-curves, ALARP principle
- Description of the method:

  According to the "Guidelines for formal safety assessment" document (IMO 2002), FSA is "is a rational and systematic process for assessing the risks relating to maritime safety and the protection of the marine environment and for evaluating the costs and benefits of IMO's options for reducing these risks". The document also states that the FSA "may be particularly relevant

for proposals for regulatory measures which have far reaching implications in terms of costs to the maritime industry or the administrative or legislative burdens which may result". FSA is intended to be used at the IMO-level, i.e. at the level of international maritime regulations. FSA puts a lot of weight on the availability of suitable data for each step, making risk acceptance criteria explicit, the format of the resulting FSA report and on Human Reliability Analysis (HRA) – even if the latter reflects classic "human error" approach of Safety-I. Different possible indices such as *cost of averting a fatality* (CAF) are introduced.

The "Amendments to the guidelines for FSA" document (IMO 2006) gives updated and more detailed guidance on the methodology. FN-diagrams are introduced in more detail, and it is stated that the society has a strong aversion to multiple-casualty accidents. The ALARP principle is also explained in more detail. It is recognized that "societal risk acceptance criteria cannot be simply transferred from one industrial activity to another". Risk is defined to be intolerable if it exceeds the average acceptable risk by more than one order of magnitude, and it is negligible (broadly acceptable) if it is one order of magnitude below the average acceptable risk. These upper and lower bounds represent the ALARP region. Examples of risk acceptance criteria are also given in a very concrete way. Interestingly, the document also recognizes the need to perform "a qualitative evaluation of risk control option interdependencies", which is a step towards recognizing the interconnected nature of problems (with other problems) and solutions (with other solutions and problems).

The Rail safety and standards board (RSSB) of the UK developed the *Safety Risk Model (SRM)* which was originally published in 2001. The principal sources are RSSB (2009, 2014a, b) and Taig & Hunt (2012).

- Steps covered: input of operational incidents, updating of existing statistical risk model, extraction of wanted risk estimates.
- Inputs: operational safety data (on a yearly basis)
- Outputs: risk values for different parts of the rail operation in the UK.
- Relationship with new risk perspectives: the model is based on expected values.
- Relationship with complexity: none.
- Risk acceptance approach: no direct application.
- Principal elements of the method: complicated set of fault and event tree models.
- Description of the method:

The Safety Risk Model consists of a series of fault tree and event tree models. They represent 131 Hazardous Events (HE). These pre-defined events have been chosen to represent the total safety of the railway system, i.e. the model is able to estimate the risks of these events and the total risk as the sum of the individual risks. What is estimated is the risk taking into account the current risk controls – i.e. the residual risk. This risk is expressed in units of average number of *fatalities and weighted injuries* (FWIs) that could occur per year on the railway. The base for the calculation is simply the expected value: [expected frequency per year] times [expected FWI if event occurs]. The values in the model are based on statistics of past occurrences. There is a very clear definition on what is in/out of the SRM scope (RSSB 2009, p. 15).

On one hand, the model is very powerful, as it covers the whole UK railway system and is able to provide quantified risk information on almost any aspect of the operation. For example, an FN-curve can be produced for the whole operation (RSSB 2014b, pp. 39-43). The model is fed by 75,000 records per year (RSSB 2014a, p. 10). On the other hand, it is fundamentally limited, because it only reflects a causality defined in fixed fault/event trees and only recognizes the predefined 131 event types. It would not cover new phenomena or unexpected interactions between different hazardous event types. Assumptions (including those of the strength of knowledge) are fixed and cannot be reviewed or adjusted case-by-case. The SRM is fundamentally one big model, which is updated periodically with new statistical data. It also

does not cover the risk *management* aspect, i.e. *what* should be done to improve the situation, and *how*, but it aims at providing supporting risk information. The actual risk management process would need to exist in parallel to the SRM.

Australia and New Zealand have been leading countries in risk management and safety management. One indication of this is that the current international standard on risk management – ISO 31000 – was developed based on the AS/NZS standard. Before reviewing the Regulatory Safety Management Program of the Australian Civil Aviation Safety Authority, it is useful to review the ISO 31000 standard for risk management, because the former is based on the latter.

The *ISO 31000 Risk management principles and guidelines* by the International Organization for Standardization (ISO 2009) is the recognized international standard for risk management.

- Steps covered: Establishing the context, risk identification, risk analysis, risk evaluation, risk treatment, communication and consultation, monitoring and review.
- Inputs: identified risks (using various techniques and tools).
- Outputs: results from risk assessment, risk treatment options.
- Relationship with new risk perspectives: the definition of risk relates to *uncertainty*, and the need to consider the *confidence* in the determined level of risk and assumptions in mentioned. However, the approach within the standard is otherwise fairly classical, and does not make explicit provisions for black swans, for example.
- Relationship with complexity: none.
- Risk acceptance approach: only broad guidance is given.
- Principal elements of the method: risk management principles, framework and process.
- Description of the method:

  The standard is not a risk management method in itself, but rather a collection of guidance for setting up a risk management framework and process. The terminology is clear and well-defined and creates a logical set of concepts. The three main elements of the standard are the principles, the framework, and the process. The principles give very high-level guidance, such as "being transparent and inclusive". The framework is about designing, implementing and improving the framework continually. Finally, the process defines the flow of the risk management activity in more detail and especially its visual presentation clarifies the relationship between different terms nicely. The documentation explains each step in the process to some extent but without going into details and without specifying exactly how the implementation should be done. Overall, the standard is a valuable high-level guidance material which can be used to make sure that all key steps in the process have been taken into account and that the guidance provided has been taken into consideration. The ISO 31000 risk management process is presented in Figure 16.

The *Regulatory Safety Management Program* (Australian Civil Aviation Safety Authority, CASA, 2015) documents the internal management program used by CASA to conduct its aviation safety activities. Safety risk management is embedded in this program but seen as a shared responsibility between industry and government aviation agencies. CASA's role is to ensure that the industry is appropriately managing the risks associated with its activities, and, failing that, take appropriate action. CASA's risk management construct is based on the ISO 31000 standard.

- Steps covered: hazard identification, risk identification, development of standards and guidance material, assessment of applications against standards, issuing of authorizations, conducting of surveillance on industry Authorization Holders, enforcement, analyzing surveillance results and other safety data, determining safety performance of Australian industry, safety education and promotion, reviewing international developments.
- Inputs: safety issue reports, investigation reports, safety-related data.
- Outputs: safety plan, sector risk profiles, industry risk profile.
- Relationship with new risk perspectives: none.

- Relationship with complexity: none but stakeholders identified very widely.
- Risk acceptance approach: acceptable level of safety performance (ALoSP)
- Principal elements of the method: full risk management process embedded within the state safety management program.
- Description of the method:

The risk management process is embedded in the "safety output process". The information processing takes place in a series of meetings conducted at different organizational levels of CASA. Risk management is split to the following levels: regulatory, surveillance, sector profiles, industry profile, system profile, and the CASA safety plan. The safety plan covers three years but is updated annually. Sector risk profiles feature systematic safety risk picture, a risk register and a high-level risk management summary. For each sector, the stakeholders are identified with a very wide scope, and at least some of them are involved in the process (CASA, 2014). Decisions are supposed to be evidence-driven.

The available documentation does not offer full detail on how the various sub-processes are carried out, but it is clear that the process is very comprehensive and reflects the ISO 31000 standard.

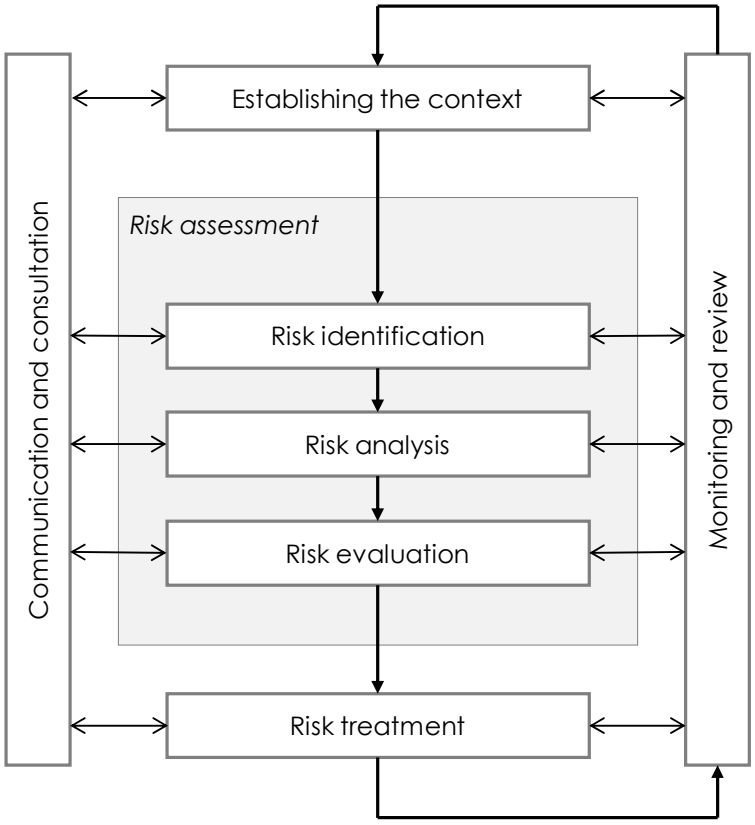


Figure 16. The risk management process according to the ISO 31000 standard. According to Standards Australia AS/NZS ISO 31000:2009 Risk management principles and guidelines (2009)

The *ARMS methodology for operational risk assessment in aviation organizations* (ARMS 2010) was created 2007-2010 by an industry working group, where the author was also involved in. The working group emerged spontaneously to create a methodology with the practical guidance for risk assessment, lacking from the ICAO safety management manual. The participants were mainly people from airlines

involved in safety and risk management activities. The name ARMS stands for aviation risk management solutions.

- Steps covered: initial risk and urgency assessment of safety events, risk assessment of safety issues including tolerability.
- Inputs: safety events or safety issues (to be risk assessed)
- Outputs: risk index and urgency values for events and acceptability for safety issues
- Relationship with new risk perspectives: none.
- Relationship with complexity: none.
- Risk acceptance approach: classic, based on numerical limits.
- Principal elements of the method: Event Risk Classification (ERC) and Safety Issue Risk Assessment (SIRA).
  Description of the method:

  Above all, ARMS is rooted in the pragmatic reality of airline safety work, with a constant flow of safety events and the need to scan through them quickly and spot issues which need urgent action, while also getting all events risk assessed for future use. ARMS delivered a clear conceptual framework for dealing with safety events and issues, in contrast to a lot of confusion around these topics at the time. Every incoming safety event is first submitted to event risk classification (ERC), which is a quick initial assessment on how much risk *was* associated with a historical safety event. The two criteria are: what would have been the most credible accident outcome *if* the scenario had escalated into an accident; and what is the probability that this escalation would have taken place. The assessment is done based on the available incomplete information because the first assessment is needed quickly. As a result, the analyst gets a level of urgency, and a risk index. The fundamentally important aspect of the event risk classification is that it focuses only on a single event and thus does not consider the number (or frequency) of events in any way. Consequently, the risk index values could be summed together across different events to get an idea of the accumulated risk value related to a certain part of the operation.

  According to ARMS, analysis of the event data will help identify *safety issues*, which then need to be risk-assessed separately with the safety issue risk assessment (SIRA). SIRA helps in defining the risk assessment task properly and in carrying it out, step-by-step. At the end, the combination of severity and probability is compared with the acceptable risk values and the acceptability of the risk is obtained. Both ERC and SIRA are based on a very simple linear model reflecting the Reason model. Probability assessment is based on barrier analysis. The methodology is well-documented and in addition to the main 67-page document, there are several additional documents, including the flowchart which presents the full risk assessment process as defined by the working group (see appendix 1) and an example SIRA excel tool (see appendix 2). The documents are available at: http://skybrary.aero/bookshelf/content/index.php

  ARMS was adopted by a large number of airlines and they use the accumulated ERC value as a proxy for current operational risk level. ARMS offered an alternative to simple safety indicators which are based on *counting* different types of events without any consideration for the potential severity of these events. Nisula (2014a) argues that ARMS has an advantage compared to the safety indicator approach. The methodology has also been the basis for further developments of similar methodologies, including the recent European (EASA) standard for risk assessing aviation safety events.

  ARMS provides a good basis especially for the initial risk assessment of events (ERC) and it is valuable that the large adoption of the methodology shows its validity in the real operational environment. However, as such it is only suitable for aviation operations of fairly large aircraft and is sensitive only to safety risks.

It is interesting to note the link between Safety Issue Risk Assessment (SIRA) in ARMS and the risk assessment of changes, required in SMS frameworks, as seen in Chapter 4.1.3. Both take place at the same conceptual level, i.e. not at the level of individual events but at the level of "an issue", something aggregated from several smaller factors. In both cases, the issue can be defined and scoped in detail before the risk assessment, which helps the risk assessment a lot and enables more consistent results. In both types of risk assessments, the objective is to understand whether the risk is acceptable as such, and if not, what could make the risk decrease to an acceptable level. In other words, a Safety Case is created (analogous to a business case). The similarity means that SIRA may be a useful tool for risk assessing changes. In any case, it is clear that there is a need for risk assessments at this level both for the transport operators and at the agency level.

Keeping in mind the refined requirements for risk management frameworks, it is easy to see the shortcomings in the current practices. Taking a step back, one can ask how such shortcomings affect the ultimate task of safety management. At this level, one can immediately recognize at least the following problems:

- Risk identification and analysis are incomplete due to lacking emphasis on knowledge-building and black swan scenarios.
- Prioritization of risks is handicapped because the knowledge dimension is missing. Above all, risks with low probabilities combined with low knowledge levels will not get the attention they deserve.
- Especially in the context of transport operations with very high safety levels, there is no practical way to deal with the acceptance of operational risks based on some numerical hard limits. Safety authorities cannot specify meaningful acceptable risk limits for operators, and even if they tried, operational risk levels cannot be measured in order to make the comparison.
- There is no practical way to implement priority-setting such as the one of EASA (see Chapter 5.3) so that the priorities would influence decision-making systematically.
- The lack of a holistic risk picture with benchmarks of existing risk levels, and the mentioned difficulty in dealing with risk acceptance, leave room for case-by-case decisions influenced too easily by individual views, including those of politicians.
- Risk treatment has not developed to the stage where it could work effectively within a complex system.

In summary, some safety threats are missed, risk assessment is often skewed, and safety interventions are not adapted to the complexity of the real world. Additionally, setting priorities and acceptable safety levels is problematic, and this leaves too much room for decisions based on individual views.

4 Existing scientific frameworks for risk management

This section focuses on risk management frameworks and related methods available in scientific literature. Some of the methods cover the risk management process quite comprehensively, while some only cover a part.

Risk, surprises and black swans by Aven (2014) contains a number of recommended principles, approaches and methods for risk management, also introduced in several scientific papers (e.g. Flage & Aven 2009, Aven 2013a, b, Aven & Krohn 2014). These methods are considered here as components of one overall process, thereby making it a fairly comprehensive framework.

- Steps covered: identification of failure scenarios, risk assessment, risk acceptability, managerial review, decision-making.
- Inputs: given activity with its objectives
- Outputs: comprehensive risk assessment and decision-making.
- Relationship with new risk perspectives: new risk perspectives are fully considered.
- Relationship with complexity: adaptive risk analysis is discussed; otherwise not explicitly covered.

- Risk acceptance approach: clear procedure proposed, reflecting the importance of SoK, in line with the new risk perspective.
- Principal elements of the method: risk assessment taking into account the assumptions and the underlying SoK, addressing surprises/black swans, judgments on risk acceptability and relative benefit of various measures reflecting SoK, managerial review & decision making.
- Description of the method:

  The starting point is the identification of failure scenarios. The two featured methods (anticipatory failure determination, red teaming) can be applied in a wide range of activities but are as such not specifically designed for in-flowing safety data. Different ways to address the SoK-dimension in semi-quantitative risk analysis are proposed:
  - Direct grading of SoK using a simple (strong-medium-weak) scoring, based on Flage & Aven (2009), see Chapter 2.2 for the scoring criteria. This implies a single SoK value over the set of assumptions underlying the risk assessment.
  - Treating each assumption separately and determining the assumption deviation risk for each assumption. A rough, simple way of doing this is only to consider the SoK. This can then be integrated with the sensitivity analysis of the probabilistic result.
  - Using the assumption deviation risk and considering for each potential deviation the magnitude and its consequences, the uncertainty and the background knowledge. If the SoK is not strong, the mini-risk assessment for the particular assumption deviation can be upgraded. The judgment on the overall SoK behind the probabilistic risk assessment can then be based on the individual assumption deviation risk results.

  Black swan risks are addressed in the risk assessment by making a list of potential black swan events. Several ways are proposed for making sure that the list becomes as comprehensive as possible: using different groups of people, considering low-probability scenarios from the risk assessment, as well as historical events. All this information is communicated along with the core risk assessment. Adaptive risk analysis and robust analysis are also introduced.

  The risk management strategy is a mix of risk-informed risk management, robustness and resilience-based approaches and discursive strategies. The proposed risk acceptance method integrates the SoK dimention: e.g. even a risk which would be acceptable with large margins based on a probabilistic approach, may become unacceptable if the SoK is weak. Similarly, in decision-making on specific measures, their impact on robustness and resilience, as well as the underlying SoK are taken into account. Robustness, resilience, discourse-based approaches and focus on SoK are recommended in addressing risks related to surprises.

  Overall, the methods introduced in this source follow perfectly the principles associated with the new risk perspectives. Tangible examples are offered on how the SoK and potential surprises can be taken into account in the risk assessment and while pondering alternative risk treatment measures.

Leveson (2015) proposes *a systems approach to risk management through leading safety indicators*.

- Steps covered: hazard identification, monitoring validity of pre-defined assumptions and feeding results to the risk management program.
- Inputs: the design and operation specification of the system in question.
- Outputs: safety-critical assumptions and their validity monitored over time.
- Relationship with new risk perspectives: none.
- Relationship with complexity: partly based on STAMP, which considers interactions between system components and dynamic adaptations within the system. The scope may not be the entire system, but it is fairly large, including the regulators and law makers, for example. The nature of CAS is not fully embraced, e.g. in terms of interventions and unintended consequences.
- Risk acceptance approach: not covered.

- Principal elements of the method: identification of hazards, scenarios and safety-critical assumptions. Creating leading indicators and "embedding them within a risk management program". Monitoring validity of assumptions and/or effectiveness of prescribed actions.
- Description of the method:

  The basic hypothesis behind this method is that "useful leading indicators can be identified based on the assumptions underlying our safety engineering practices and on the vulnerability of those assumptions rather than on likelihood of loss events". Rather than modeling the system in a probabilistic manner, the system specifications are used to create a set of safety critical assumptions. These assumptions are transformed into leading indicators which are monitored, and for which predefined action plans exist for cases where the indicators would show that the assumptions are no longer valid. For complex systems like aviation, it can be expected that the number of assumptions and indicators would be quite high, e.g. several hundreds. It is stated that the leading indicator program should be embedded in the risk management program but how this should be done has not been communicated.

  Because this approach is not a risk management process per se, it does not deliver any risk values. If such values are required, then this method is probably not suitable. It is also difficult to estimate the resources required first to create the framework of hazards, assumptions and indicators, and then to monitor and action these items operationally.

Nisula (2015a, b) introduced an *integrated risk picture for four modes of transport* and adapts the ARMS method to *integrate safety events in the risk management process*. As a part of the approach, he introduces the so-called *safety factors* (Nisula 2014b).

- Steps covered: capture of safety data and information, knowledge integration and building, creation of risk picture, defining interventions with the transport system.
- Inputs: all safety data and information available in the state transport safety agency (safety events, audit reports, knowledge of experts, etc.)
- Outputs: risk picture, safety interventions.
- Relationship with new risk perspectives: the new risk perspective with uncertainty and black swans is embraced.
- Relationship with complexity: the transport system is considered a CAS.
- Risk acceptance approach: not explicitly covered.
- Principal elements of the method: integrated risk picture for 4 modes of transport, including event risk data.
  Description of the method:

  The method aims at embracing the new risk perspectives, as well as the understanding of the transport system as a complex adaptive system, and strike a balance between the scientific requirements and the more pragmatic operational requirements, typically related to time and resource available. The ARMS methodology is adapted and used to transform the operational safety events into risk information. Threats and scenarios from four modes of transport, as well as the safety event data, are all integrated in one risk picture. This picture is the basis for comparing different risks and taking decisions. Interventions with the transport system reflect the CAS nature of the system.

  In the development of this methodology, a new concept emerged. Because the event risk assessment process treats both incidents with only potential outcomes and also events with tangible actual outcomes, it would not be correct to talk about risk (which is associated with potential outcomes only). The new concept, *riss*, is a fusion of risk and loss. It is explained in more detail in Part II. In general, the methodology presented in the mentioned articles by Nisula was the prototype of the NRMF introduced in this dissertation. Therefore, the details of this approach will be covered in Part II.

5 Synthesis and conclusions

The first section provides a lot of useful information for different aspects of risk management and the related interventions. There are a few key themes which are reflected in more than one source. There is agreement that *risk management needs to be integrated with the overall decision-making process* and take into account all the *different priorities, not only safety*. Creating a *holistic risk picture* is recommendable and if necessary it could be only semi-quantitative. Decisions and judgments cannot be left to algorithms but need to be *made by humans* who can try to integrate together all the different priorities and value judgments. *Stakeholders* need to be involved to give their inputs and feedback. An *adaptive approach* needs to be adopted for risk management and interventions. This means generating a lot of different alternatives and testing them, preferably in small-scale, before gradual larger implementation. Concrete guidance for adaptive policymaking has been provided. Postponing the recovery of additional information for decision-making can also be considered part of the adaptive approach, if it allows early testing to start immediately.

The more detailed advice presented in the various lists should be useful as such, including the recommended phases for the risk management and decision-making processes, challenges related to risk matrices (Duijm), and Swanson's points on adaptive interventions and policies.

Some important challenges are mentioned, such as the ability to *communicate* the understanding around risk analyses and proposed alternatives properly *to the decision-makers*, communicate *with stakeholders* on the phenomena involving high level of complexity, and *gaining acceptance for increased decentralization and self-organization* despite the increasing uncertainty and apparent loss of centralized control.

The proposed approaches seem to fit together well. In particular, the proposed decision framework (Aven et al. 2007) puts a lot of weight on generating different alternatives and points out that risk acceptance is more about alternatives that risks themselves; the framework also promotes a better visualization of the different trade-offs related to decisions. This fits very well with the idea of creating a risk picture with semi-quantitative inputs and allowing a comparison of the different alternatives (Aven 2008).

The review of the existing risk management frameworks shows that none of the current frameworks fully integrate the desired features: modern understanding of risk and safety (II), as well as understanding the surrounding system as a complex adaptive system and adapting the interventions accordingly. Having said that, many of the related frameworks contain elements of these modern approaches.

The reviewed methodologies from the industry are very traditional in their approach to risk, even if the CASA program has embraced the ISO 31000 world and thereby contains many positive elements like the reference to uncertainty and the consultation of the stakeholders. The industry methods stress the need to absorb operational safety data as an input to the process and have acknowledged the role of the data sources and the hazard identification process. However, as summarized above, the failure to embrace the modern risk perspectives and complexity introduces major shortcomings at every stage of the risk management process.

The scientific methods are less sensitive to the need to handle large amounts of operational safety data as an input but are much more advanced in their approach to risk assessment and decision making. Not surprisingly, the approach in Aven (2014) reflects perfectly the new risk perspectives and the associated modern approach to risk management. None of the reviewed methods covers the whole process from the beginning to the end, i.e. from the operational safety data to the interventions with the system, using approaches adapted to complexity.

Structurally the reviewed methods feature two different types: some are based on a large operational safety model, possibly containing sub-models made of fault or event trees (RSSB, Leveson) - while other methods can accept almost any type of data/information as input (ICAO, ARMS, Nisula). The IMO method (as well as many other methods not reviewed in detail) are made for a single detailed assessment instead of a continuous operational risk management activity. As one of the basic starting points in this dissertation is the notion of a complex adaptive system, it is difficult to imagine using a preconceived operational model with a limited set of predefined accident outcomes (and often also predefined causal mechanisms) as the basis for risk management. It is hard to believe that such an approach could reflect well enough the dynamic reality which is changing the operational system and its risks organically all the time. *It is considered preferable that new hazards and risks can be identified in a flexible way at any point in time and that phenomena like black swans, which would never be part of a model, can also be addressed within the risk management process*. The risk management framework needs to be able to *function in pace with the real operational world* and digest the continuous flow of operational safety data, as well as the continuous interactions within the transport system.

Conclusions of Part I

The purpose of Part I has been to review all the key topics supporting the creation of a risk management framework adapted for the transport system and reflecting the state-of-the-art scientific understanding of the contributing subject matters. In the end, available current risk management frameworks have also been reviewed.

The three key topics which emerge from this study are the *modern risk perspectives*, *complex adaptive systems* and *resilience engineering*. These three approaches coming from quite different domains fit together well and have a significant synergy.

The modern risk perspectives stress the importance of *uncertainty*, rather than use the usual probability dimension, paired with severity. Uncertainty is also a fundamental property of complex adaptive systems. Uncertainties in such systems are huge and often fall in the domain of *unknowable* unknowns. The complexity of these systems also set the scene for *black swan* events, which are another aspect of the modern risk perspectives. Due to the uncertainties and the very high number of potential scenarios within complex systems it is impossible to address all scenarios one by one, from the safety management perspective. This is where resilience and resilience engineering come in the picture: the idea is to build generic resilience which will address a large number of different threats and scenarios simultaneously.

Because the organizations within the transport system are complex adaptive systems themselves, both safety and resilience are *emergent properties* and cannot be controlled directly by anyone, including the top management. The situation for safety, risks and resilience within the system is very dynamic and as the knowledge and information are spread around the system, it is already challenging to be aware of these system properties at any point of time.

The study of these three main topics also reveals that for effective risk management, the process needs to address the system at three different levels. Threats, incidents and scenarios belong to the *level of events*. Managing these and the emergence of safety and resilience belong to the *organizational level*. Trying to interact with the transport system and introducing positive change successfully will only be possible by adopting a perspective which looks at the transport system at a *third level: as a complex adaptive system*.

The adopted risk acceptance approach abandons simple numerical criteria and promotes a *holistic approach where the many dimensions of risk are taken into account* and where comparisons to absolute reference values are replaced by comparisons to other risks, focus on the whole decision-making process and comparisons between different risk treatment options, taking into account both the risk reduction and the various costs involved.

Despite the many illustrated limitations and biases of humans related to risk perception and decision making, the adopted approach still *places the humans in the central role within risk management*. This is due to the irreplaceable capabilities of humans to integrate very various kinds of aspects and perspectives together, understand the relevance of (historical, psychological, emotional, etc.) factors affecting human behavior, transfer information within its context to other humans and to make ethical judgments.

The key findings and adopted paradigms on each subject matter in Part I have been summarized in the conclusions at the end of each chapter. Importantly, these findings have also led to a number of *requirements for the NRMF*.

It becomes clear in Chapter 3 that the *conditions for controlled scientific experimentation cannot be fulfilled when the object of the study is a complex adaptive system*. By definition, such systems change organically continuously, and cause-effect relations do not exist in the classic sense due to the multiple

simultaneous factors influencing the system and due to the numerous interactions and feedback-loops. This has important *implications also for validating the usefulness of the developed risk management framework*: simply testing the framework in real organizations cannot be used to prove its value nor its lack of value - even if the experiences may give important hints on parts which seem to work well and parts which seem challenging. Therefore, the importance of the derived requirements (for the NRMF) is high also because they can be used as an important means for validating the value of the developed NRMF, as will be seen in Part III.

The research findings from Chapter 2 to Chapter 5 illustrate that there is an important body of knowledge which can be very beneficial in the context of risk management. However, the review of existing risk management frameworks in Chapter 6 shows that there is indeed *a lot of room for improvement in the way that such research has been implemented in risk management frameworks today*. The failure to embrace the modern risk perspectives and complexity introduces major shortcomings at every stage of the risk management process. From the point of view of being able to manage risks/safety properly, these shortcomings mean that some safety threats are missed, risk assessment is often skewed, and safety interventions are not adapted to the complexity of the real world. Also, setting priorities and acceptable safety levels is problematic, and this leaves too much room for decisions based on individual views. Some of the scientific methods fully embrace the new risk perspectives but are not adapted to the continuous flow of operational data as the main input. The techniques for *adaptive interventions* are generally not part of risk management frameworks. Both the review of research related to complex adaptive systems in Chapter 3 and the discussion on adaptive policies in Chapter 6.1 highlight the importance of adopting intervention approaches which are adapted for complex systems.

The starting point for the current work was the postulated mismatch between today's risk management methods and the existing scientific knowledge. The conclusions of Part I support the existence of this mismatch and therefore the usefulness of the undertaken effort.

Based on these conclusions and the detailed requirements, a NRMF will be developed. The existing shortcomings will have to be addressed in different ways: comprehensive risk assessment based on the new risk perspectives helps capture black swans and other risks which are missed with current methods and produces a robust result incorporating the knowledge-dimension; a practicable but scientifically robust method for risk evaluation needs to be proposed, based on alternatives and benchmarks rather than hard numerical limits; a holistic view to all risks needs to be provided through a risk picture; adaptive risk management approaches need to be proposed; and finally the methods need to be implementable in a real operational set-up where key steps in the process need to have short lead times and the overall process needs to be effective in highlighting the top risks and interventions without excessive spending of resources.

PART II – A Risk Management Framework for a Complex Adaptive Transport System

Chapter 7 – Modern Risk Management Framework for four modes of transport

This Chapter introduces the developed Risk Management Framework. The content is organized according to the ISO 31000 standard. Reference is made to the requirements developed in Part I.

The various components of the framework are first discussed in detail, one by one. They are then reviewed as a process in Chapter 7.6. The Chapter ends with a discussion and conclusions on the developed Risk Management Framework. The reader is reminded that the case study of Chapter 9 contains examples of the implementation of the framework.

1 Setting the context

The context for the NRMF is the role of a national transport safety agency, having four transport modes in its scope, and having identified other important priorities besides safety, that need to be addressed simultaneously. The specific context and objectives of the Finnish Transport Safety Agency, Trafi, are reviewed under the case study in Part III.

The NRMF must be such that it is feasible to implement it in the real-life circumstances (which is also reflected in the requirements). Its main objective is to guide the Agency in using its resource in the best possible way to minimize the safety risks in the transport system, while respecting other key priorities, such as environmental sustainability and reliability of transport services. Ultimately, pursuing this objective has to be reflected in the decision-making on interventions and tolerability of risks.

The NRMF must cover the whole process, starting from the reception of operational safety data and ending with interventions with the system and monitoring the obtained results and the system performance in general, and feeding these back to the beginning of the next cycle.

Safety agencies are supposed to oversee the whole transport system and to provide additional safeguards for the safety of individual citizens using the transport system. The scope of the agencies is therefore exceptionally large, covering all the operators (typically within some geographical or political boundaries) and a large number of citizens engaged in hobbies also within the scope of the agencies: e.g. sailing or operating drones. Safety agencies are also part of the national political system and international standardization activities to be carried out through international organizations such as the International Maritime Organization (IMO) and the International Civil Aviation Organization (ICAO). There are requests for studies, questions and directions coming to the agency from these bodies, often related to legislative processes and therefore “must do” actions by given deadlines.

Due to these factors, it seems the national safety agencies have more requirements and constraints for their risk management framework than other organizations within the system. In some cases, the agencies overseeing different modes of transport have been grouped together as one agency, which is the case in the case study, for the Finnish agency. This arrangement may produce synergies but may also prove to be particularly challenging if several modes of transport are to be managed with a single risk management framework. Therefore, in developing this NRMF*, a national multi-modal safety agency is taken as a reference user*. Consequently, it is assumed that the resulting framework should be useful at least for any organization within the transport system.

2 Risk identification

The requirements for risk identification are:

- (RI-1) As the transport system is constantly evolving, and so are the risks within, the NRMF needs to include provisions to feed in a continual flow of data/information from the system.
- (RI-2) There is already an established flow of *safety information* within the transport system. The NRMF needs to be able to digest *the types of safety information* used currently and transform such data into useful risk information.
- (RI-3) The need to carry out a risk assessment related to a safety issue or a change must be one of the normal inputs to the process.

The NRMF is designed to digest three basic types of safety data/information:

- Flow of *operational safety data*, which results in:
  - Safety events (in their raw format). Such reporting to the agency is often mandated by law, e.g. "mandatory occurrence reports" in aviation.
  - Safety Issues
    Information which can be obtained from the operations directly in the form of analysis results, rather than raw data (e.g. Flight Data Analysis statistics).
- Data/information originating from the agency's *oversight activity*, such as:
  - Audit findings and audit reports
  - Any associated information/knowledge
- *Information residing within the agency* and developing on a longer time frame. Typically:
  - This type of information comes most naturally from people participating in the process. It can be partly tacit knowledge, which can only be obtained within the right context, i.e. during a discussion about a specific topic.
  - Its origin is often in the experts' own personal experience, including professional activities before joining the agency, and enriched by the activity at the agency and the information available at the agency.
  - Safety research can contribute to such knowledge (e.g. accident investigation reports and the numerous studies done on various aspects of road safety).
  - In addition to identifying the *current threats*, the experts' knowledge can be specifically targeted to identify *future threats* (threats which do not exist yet, at least at a significant level, but are expected to emerge in the foreseeable future due to various trends). This promotes anticipation of risks.

The most visible type of operational safety data for a safety agency today are the safety event reports, due to their mandatory status and long traditions, especially in aviation. Such reports are narratives, reported on standardized reporting forms, with standardized fields. Other types of safety data may be at least as important for the operators themselves, like flight data analysis (FDA), where the raw data is electronic recordings of hundreds of parameters on commercial flights, but the results can be presented in the form of statistics, and such statistics may be available also for safety agencies. FDA data is not available to safety agencies in their raw format due to confidentiality constraints.

Safety Issues are identified "potential safety problems". A safety issue may be identified due to aggregation of safety events (e.g. a pattern of similar events, see ARMS process in Appendix 1), or directly, e.g. based on a single event leading to a specific concern, or cascaded from the ministry of transport. The need to build a safety case (e.g. on a proposed change) leads to the same situation where the input to the process is a safety issue.

With the mentioned three types of safety data/information as inputs, and adding that also experts from outside the agency are meant to take part in the process, the requirement (RI-1) can be fulfilled. As the process can digest safety event/issue data, safety information in a report format, as well as information/knowledge directly from people, the requirement (RI-2) is also fulfilled, at least for its data input part. The part concerning the transformation of such data into risk information is covered under

Chapter 7.2.1 for safety events and under Chapter 7.3.3 for safety issues – so requirement (RI-3) is equally addressed.

The specific case of organizational safety profiling is handled apart in Chapter 7.5.6. Such data is not used primarily for identifying threats, but rather for assessing the organizational resilience.

In practice, the current safety analysis approach is to use the raw data to build various kinds of analyses. Events are typically categorized in multiple ways and statistics can be created based on these categories. Such statistics can be used to identify safety issues, and in today's world the issues identified this way would be reviewed in some kind of expert or management team, and if considered necessary, relevant actions would be launched. The identified safety issues could also be called risks, whether they would be actually assessed or not. Safety indicators are also part of these preliminary analyses. They are fundamentally only counting how many times different events or factors occurred in a certain time period or certain operation, etc. For a long time, safety actions have been guided purely on such preliminary results based on categories and statistics.

The developed NRMF considers these results only preliminary and carries on in three different ways. First, these preliminary results can be used as one input to the risk analysis. Secondly, the identified safety issues enter the risk assessment process at their own level, as mentioned above. Finally, the raw safety event data as such needs to be transformed into risk information before it can enter the main risk analysis and evaluation process. This transformation takes place through event risk assessment. Due to the preliminary nature of such risk assessment, the topic is covered under "risk identification" despite it being already a type of risk assessment.

2.1 Event Risk Assessment

The first real step in the risk management process is the event risk assessment. It can be applied to all safety data which describes individual events. Like in the original ARMS method, the objective is twofold: first, to understand what was the risk involved in a specific historical event; and second, being able to treat a large number of events through their cumulated *risk* rather than only counting numbers of events. As the context of ARMS was an airline, learning about the urgency of the event was of primary interest: an airline safety office is the first place where a fresh safety event would arrive, and in an extreme case, the event might highlight a threat which requires stopping a part of the operation immediately, e.g. grounding a particular aircraft. The resulting colour in the original ERC matrix gave the answer regarding the urgency. The urgency aspect is less of a concern for the agency who in any case receives the report with some delay. The primary purpose for the agency is risk identification by aggregating together the event data in a way, which reflects also the potential consequences, and not only the number of events. Event risk assessment gives each event an event risk value which reflects both the *type of accident* which could have resulted (irrespective of how "far" from this accident the event stopped) and the *estimated likelihood* that the event would have escalated to that accident. Summing together the event risk values from different events gives cumulative event risk value which can be very useful in identifying threats, be it for a specific part of the operation, specific time of the day, specific location, etc. It must be stressed that the event risk assessment is based on an initial rough estimation of risk, not least because the raw event reports are often very brief, and the information may not be fully reliable. However, as a whole such a data set does point to the significant differences of risk between different parts of the operations and is thus a good approach for risk identification. The whole event risk assessment is only one input to the main risk assessment process which takes place at the level of threats and safety issues (see Chapter 7.3.2).

The requirement (C-1) is very important in the context of this step because the safety events can be very numerous, adding up to several thousand reports a year, and therefore each minute spent on risk assessing a single event adds up to a significant workload on the yearly basis. Therefore, the way to carry out event risk assessment is very much a compromise between the need to carry out a good assessment, and the need to do it fast. This constraint was very clearly part of the original ARMS work

and explains the very simple model behind the event risk assessment method. It was decided to use the ARMS ERC method as the basis but customize it to the 4-mode transport system.

Expanding the original ARMS methodology to a much wider scope than an airline introduces new challenges. The severity scale in ARMS ERC was customized for airline operation. The highest severity category is very "wide" (three or more fatalities) but in practice limited by the size of aircraft in operation which brings it to a maximum of about 550 people (the typical average being much less due to smaller aircraft and passenger load factors below 100%). Expanding the ERC to several modes of transport *changes both the maximum value and the natural severity categories* emerging from different types of vehicles, ships and aircraft. A typical glider only seats one person. There may be millions of private cars operated in a single country and their maximum capacity is typically 5 to 7 people. Different kinds of sailing boats may have different sizes of crews. Trains could seat more passengers than aircraft but even bad collisions would not necessarily lead to losing all passengers. At the highest end, leisure ships may accommodate literally thousands of people and it is not unrealistic to imagine accident scenarios where all or nearly all lives could be in danger.

Another challenge is to take into account the *different dimensions of risk*. The original ARMS ERC featured severity classes which were defined as combinations of human loss and technical damage. Latest when maritime becomes part of the scope for risk assessment, environmental damage becomes an important dimension of severity. There are thus at least three dimensions which need to be addressed for a proper severity assessment: *human loss, environmental damage, material/financial damage*. If these dimensions are kept apart from each other, every event risk assessment would have three different results on three different scales. Comparison of different events (and different aggregated safety issues) in terms of risk would become very difficult as it would be impossible to prioritize the issues. For example, how would

- "3 fatalities + €2 million material loss + zero environmental damage" compare with
- "no fatalities + huge oil damage on a 250 km coastline + €500,000 material loss"?

The same problem would mean that it would be very difficult to place the risks in a risk picture because their position on the severity axis would be unclear or at least multidimensional. Due to this, much of the benefit of this risk assessment would be lost. If all three dimensions of severity could be presented *on a single scale*, then every event would have a total risk value which would be the sum of the individual risk values for the various severity dimensions. Getting these different dimensions on a single scale, however, requires *taking decisions on the relative importance of the different severity dimensions compared to each other*. For example, in a very blunt way one could ask: how big should an environmental damage be so that it would become equal with losing five human lives? Such decisions are pure value judgments. Due to this, any attempt to create relative valuation between the different severity dimensions would only create a *snapshot within a limited population of people*. Furthermore, if average values within a population were used, the result could be quite different from some of the individual values within the population. The obtained answers would also be sensitive to *framing issues*. For example, when thinking about human lives saved, one could ask whose lives are we talking about: seven-year-old schoolchildren or a 65-year-old man convicted multiple times of drunk driving? It becomes clear that there is no good permanent solution for creating relative values between the different severity dimensions and that even values acknowledged as temporary snapshots could be controversial and subject to debate. As mentioned, Duijm (2015) recommends not to try such aggregation of different types of risks due to the difficulty.

The problem of maintaining several severity dimensions for each event would severely handicap their aggregation as there would be no single clear risk value. The problem would become very tangible latest at the level of safety issues as it would be impossible to place them into the risk picture. The use of a single risk picture for all different risks and the ability to take decisions at this very global scale are so fundamental objectives that after considering the pros and cons, it was decided that even a rough way to bring all severities on a single scale is better than losing the ability to build the risk picture and the ability to optimize interventions based on total risk at the level of the transport system.

The solution adopted was to use the same severity scale for all different dimensions of severity. The units used are points. The *reference scale is about the number of fatalities* (in line with Chilton et al. 2002, see Chapter 5.2) and the anchor point is one fatality corresponding to 10,000 points. All other types of severity are transformed into points so that the same single scale can be used. The case study in Part III shows in more detail how the bridges between different severity dimensions were established, based on a valuation exercise among a relatively small group of people. The aim at this stage was only to demonstrate how such an exercise can be run and to get the first values to be used as the prototype values, rather than to obtain highly representative values. The exercise served its role as the proof of concept and produced the coefficients between different types of severities, e.g. financial loss versus loss of human life. As mentioned, the results of such an exercise can only be taken as a snapshot from the participating population and reaching some kind of generally accepted valuations on a long-term basis could be very difficult. However, there are several factors which reduce the problems related to this type of exercise:

- The obtained bridging coefficients can be kept apart as variables which can be updated at any point in time. When properly implemented, changing the values of the coefficients will update all the risk data accordingly and the whole process becomes reversible. This also allows carrying out sensitivity studies with different coefficients.
- As the severity values exist separately for three dimensions it is also possible to come back to a separate presentation.
- In the context of event risk assessment, the required precision level is not high. The focus is at the level of the order of magnitude.
- The experience obtained during the development of the methodology suggests that the number of fatalities is the dominating severity dimension and the role of the other dimensions is often insignificant. An exception would be environmental damages within the maritime context.
- If there is a need to consider more than three dimensions of severity the additional dimensions can be treated with the same established approach.

As all severities are measured in points and probabilities are real numbers, the resulting risk values are also expressed in points. This gives the huge advantage that *all event risks have the same unit and become comparable and can also be added together to form cumulative risk values*.

Consequently, the original ARMS event risk classification has expanded into a wider-scope multidimensional event risk assessment. The severity scale can run up to any desired value, and in principle any severity value on scale can be assigned to an event without the need to stick to predefined categories. The probability scale naturally starts with one and can run to as low probability values as desired. Figure 17 presents an example of an event risk assessment space with two example events.

This two-dimensional space shows the position of the event once the risk assessment has been done. Note the configuration where the horizontal axis runs from right to left so that the end point value 1 of the semi-open probability scale can be placed at the intersection with the vertical axis. This avoids cutting the open end of the scale at a given value. Another possible configuration would be the mirror-image where the vertical axis is on the right side.

In designing the *interface for the analyst* carrying out the event risk assessment, the challenges highlighted by Duijm (2015) became very tangible. If a matrix is used, the task becomes simple in the sense that the analyst only needs to pick one square in the matrix. However, how to choose the number of rows and columns? Having only a few categories would mean that the range covered by one category is very large which introduces the *limited resolution* problem. The severity categories should also naturally fit with the typical scenarios in the particular mode of transport, and the available precision of the information available. The fact that the available precision may be very different from one case to another does not help designing the matrix. Duijm's (2015) recommendation was a continuous probability consequence diagram - which the resulting presentation is, as seen in Figure 15 – but should the interface for the analyst also be one, despite the analysts' attraction to simple matrices?

The developed solution is a combination of a matrix and a continuous scale, trying to take benefit of the good sides of both. The analyst has the choice which entry mode is used, case-by-case. If information is

very imprecise and/or the assessment needs to be done very quickly, it may be preferable to use the matrix. However, the assessment can also be done by entering specific values for severity and probability. If at least some events information is precise, this method may be preferable.

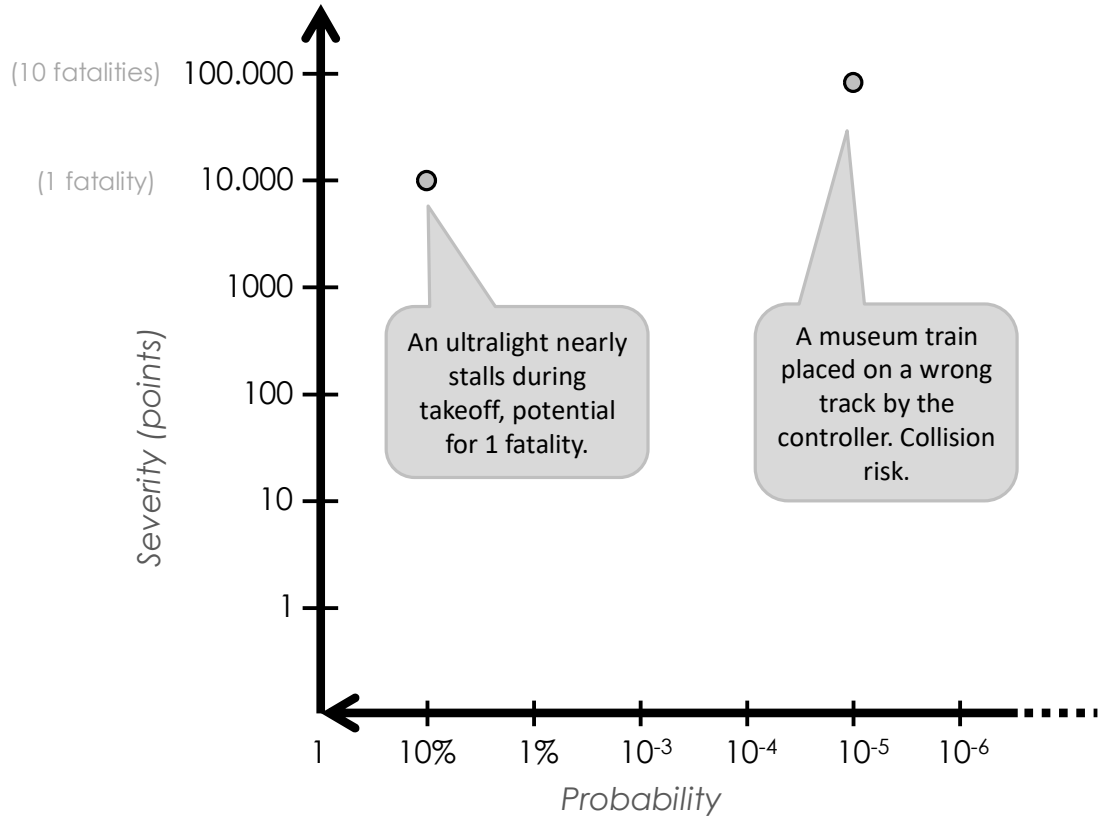


Figure 17. Example of Event Risk Assessment. Two events placed in the two-dimensional space.

To get a good fit with the natural severity categories for each mode of transport, the event risk assessment matrices were customized. Importantly, the questions to ask in relation to the severity and probability dimensions are still the same as in ARMS: first, if the event had escalated into an accident, among the possible accident outcomes, which one would have been the most plausible? The severity should reflect this scenario (and this chosen reference scenario is noted). Second, what is the probability that the event would have escalated along this chosen scenario (usually in terms of failing barriers)? The matrix for marine events is presented as an example in Figure 18. In this case, only four severity categories were used. Despite the customization, the results from each mode of transport are fully compatible and comparable due to the same reference scales both in terms of severity (the fatality scale in points) and probability. The matrix also illustrates well the limited resolution problem, particularly worrisome for the top severity category which could contain events with 100 potential fatalities as well as 5000 potential fatalities - yet this category would only be represented by a single “average“ severity value.

As an example, if the potential escalation of an event would have caused an accident where the number of fatalities can only be estimated roughly but clearly belongs to the category 5-99 fatalities, the analyst would probably use this matrix, and if the escalation probability is estimated to be about 10%, the result would be the square C52. However, in the case that the actual souls-on-board number is known to be 37, and the scenario would consider them all to perish, then it would be wiser to calculate the severity value directly 37 x 10.000p = 37.000p. With the escalation probability of 10%, the event risk value becomes 37.000p x 10% = 3700 points.

Similarly, in an event where a single pilot takes off with an ultralight aircraft and nearly stalls in demanding wind conditions, the reference scenario would be a crash following a stall, and based on the altitude (which is known), a fatality. It would make sense to value the severity to 10.000 points rather than use a matrix with a lower resolution range (e.g. 1-4 fatalities). If the probability of escalation (based on known circumstances in the specific event) is estimated at 10%, the event would be located as shown in Figure 17.

Marine event risk assessment	Zero. Accident materialized.	Marginal	Weak	Medium	Good	***Effectiveness of remaining barriers***
	100%	50%	10%	5×10^{-3}	10^{-6}	***Accident probability***
100 -	A150	B150	C150	D150	E150	
5-99	A52	B52	C52	D52	E52	
1-4	A2	B2	C2	D2	E2	
Minor injuries	A1	B1	C1	D1	E1	
Fatalities & injuries				10^{-2}	10^{-5}	

Figure 18. Example matrix for event risk assessment, customized for marine safety events.

As mentioned in Part I, there are often standard practices concerning how safety events are processed at the agency – e.g. how they are classified and there are international taxonomies (like the ECCAIRS) to be used for this purpose. Carrying out event risk assessment does not need to change these standard practices. It becomes an additional step but brings the immediate advantage that statistics related to the standard categories can now be expressed in terms of risk instead of just looking at event count. In other words, carrying out event risk assessment produces an immediate local benefit in being able to express the results of the safety events in a new way. However, the main deliverable of this stage is being able to transform information related to safety event data in a format where it can be integrated into the overall risk management process, fulfilling the requirement RI-2.

In summary, the developed method for event risk assessment is based on the ARMS ERC but largely expanded. It is able to deal with wide ranges of severities and probabilities unlike the original ERC. It is also able to deal with different dimensions of severity (this aspect is discussed in more detail in the case study of Part III). Just like ARMS ERC, the obtained event risk values remain rough estimates and the limitations of the very basic model behind this assessment are acknowledged. In other words, the Strength of Knowledge associated with the event risk assessment values is systematically low. These limitations are accepted on one hand because the event risk assessment still brings in the severity dimension (which is a great advantage compared to safety indicators) without introducing unacceptable resource requirements, and on the other hand because the safety events are only one of the inputs to the risk management process, so crosschecking with other elements will be possible.

2.2 The concept of Riss

The purpose of event risk assessment is risk identification at the level of safety issues. The key for achieving this purpose is to be able to aggregate events and their associated event risk values together,

and to study the obtained cumulative risk values. For example, the focus could be on comparing the risks in landing on particular runways.

Three types of events can contribute to such risk identification:

1. Events which did not have any *actual outcomes,* or the actual outcome was insignificant in its magnitude. The interest in such events is to assess the risk related to the *potential outcomes*. This is the most typical case for events running through event risk assessment. For example, a landing where it was challenging to stop the aircraft before the end of the runway (possibly due to the characteristics of the particular runway and the associated approach procedure).

2. Events which already escalated to the loss. In other words, there is an actual outcome which is factual and there is no longer a potential outcome. The concept of risk does not really apply here because there is no longer uncertainty about whether or not the loss will materialize and exactly what kind of loss will materialize. For example, a landing event where the aircraft cannot stop on the runway, causing a runway overrun accident with a number of injuries and severe damage.

3. Events which already escalated to some kind of loss but where one could envisage further escalation to an even higher loss. These events are thus a mixture of type 1 and type 2 events. There is a need to recognize the actual loss as such but to add a potential further escalation. For example, a low-speed runway overrun with minor injuries and some damages to the aircraft, but with a potential for a worse accident with more severe injuries, damages, and even fatalities.

From the event risk assessment point of view, all these types of events should contribute to the cumulated risk value (e.g. for the particular runway in question). Technically speaking there is no problem because materialized losses can be assessed by setting the probability to 100%, and the potential losses will have probabilities less than 100%, and all the obtained event risk values can be summed together. However, from the conceptual point of view there is a problem because some of the consequences have already materialized and in this case one should not talk about risk, but rather about a loss. The total cumulated "risk" value gets contributions both from risks and losses, and it makes a lot of sense to use all these types of results for risk identification. This gives rise to a new concept called *Riss*. It stands for "Risk and Loss" (Nisula 2015a, b). If the safety events which go through event risk assessment process include events with actual outcomes, then strictly speaking, one should talk about *event Riss assessment* and cumulated Riss values.

In real life, the events that need to be processed contain all three types of events and it makes sense to take benefit of all these events as they all contribute to building a better understanding of where the risks are. Event risk assessment gives meaningful results for all three types of events. For type 3 events, event risk is the sum of two components: the actual outcome ($p=100\%$) and the potential outcome ($p<100\%$).

For example, let's imagine the focus is on two different airfields both featuring a very active leisure aviation operation. The question is, how could the historical events experienced at these two locations be used to provide guidance on the risk levels and therefore also on the importance of risk treatment. Let's imagine the last three years event statistics show the following:

	Airfield 1	Airfield 2
Accidents (2 fatalities)	0	2
Accidents (1 fatality)	2	1
Incidents (potential outcomes):		
2 fatalities (50% probability)	7	1
1 fatality (50% probability)	2	1
1 fatality (1% probability)	4	12

It should be noted, that in real life, every event would have its own distinct potential severity and probability – for the sake of the example, only the above five types of events are considered. Without the event risk concept, the comparison could only be based on the *number* of different types of events:

	Airfield 1	Airfield 2
Accidents	2	3
Reported incidents	13	14

This is the view that *Safety Indicators* could typically provide: counting different types of events. It can be seen, that based on accident count or incident count, the airfield 2 would seem to have a riskier operation. These numbers do not take into account the severity or the potential severity of the events. For example, incidents which could have killed two people with 50% probability are equal to incidents which could have killed only a single person with 1% probability. If one wants to take into account the severities, the *event risk* concept can be used. For example, for the first line, the calculation gives: 2 x 10.000p x 50% x 7 = 70.000p. However, the event risk concept was designed for potential outcomes only, i.e. in this example it would not cover the accidents. The incidents would obtain the following event risk assessment points:

	Airfield 1	Airfield 2
2 fatalities (50% probability)	70.000p	10.000p
1 fatality (50% probability)	10.000p	5000p
1 fatality (1% probability)	400p	1200p
TOTAL:	80.400p	16.200p

Now, one could say that airfield 1 looks riskier in terms of the risk embedded in the incidents, while airfield 2 has more accidents, which can be argued to be more factual than the cumulative risk in the incidents. This means the comparison is still difficult. The *riss* concept can be used to transform all incidents and accidents into the same points, which are now riss points, as they reflect both experienced losses and incidents with risks. For example, an accident with 1 fatality would make: 1 x 10.000p x 100% = 10.000p. The following comparison can be made:

	Airfield 1	Airfield 2
Accidents (2 fatalities)	0p	40.000p
Accidents (1 fatality)	20.000p	10.000p
Riss from events (from above):	80.400p	16.200p
TOTAL:	100.400p	66.200p

While these numbers need to be considered very rough and reflecting only one perspective to risk through historical events, the advantage is that thanks to the riss concept all types of events have been transformed into the same riss currency and their total cumulative riss-values can be compared. Despite the accidents putting more weight on the side of airfield 2, the total riss score is higher for airfield 1, thanks to the event risks of the incidents.

In event risk assessment (just like in the ARMS ERC), as far as the *potential consequences* are concerned, the event is observed from a time point before the outcome of the event was known. This means that the concept of risk is still applied and the outcomes that could have materialized are treated as potential outcomes, even if from a purely factual point of view, at the time of carrying out the assessment, the event is history and does not have any uncertainty related to it. In other words, in event risk assessment one asks the question "what is the risk that the event *would have* escalated to the potential outcome?". Obviously, for the *actual consequences*, the p=100% notion implies that the knowledge about the outcome *is* taken into account and assessed as much as possible based on the facts.

2.3 Safety Factors

The so-called Safety Factors introduce another new approach. They are an attempt to move away from classic reductionism-based event classification taxonomies and work more in line with the Safety-II thinking. Safety factors can be used in many ways but one of their key application areas relates to making sense of safety events and thus support risk identification.

Safety factors are the *assumed prerequisites for safe operation*. They are characterized by three features:

- They are phrased as *positive* statements.
- As much as possible, the safety factors are defined at the level of *functions* (such as *controllability*) instead of at the level of individual devices, procedures or other operational details.
- As a result of the first two points, the list of safety factors is short and compact, in contrast to many large taxonomies in use today.

The rationale behind the use of safety factors is the idea that detailed reductionism-based taxonomies may not capture the interactions and dynamics in a living sociotechnical system due to the focus on component level details. The safety factors remain mainly at the higher level of functions so a lot of the interactions remain intact within the functions. Importantly, the fact that the safety factors are positive enables the collection of both negative *and positive* aspects of safety.

The initial set of safety factors was drafted using three guiding principles:

- The set should cover all aspects of the operation, i.e. all high-level safety critical functions should be captured.
- Overlap of safety factors should be avoided.
- The factor needs to be a positive function, not a technical device or a failure condition.

An example set of safety factors is presented in Appendix 3. In practice, each mode of transport needs its own set of safety factors, even if there is a lot of similarity between the safety factors from different transport modes. Details on how the safety factor lists were created in cooperation with Trafi can be found in Chapter 9.2.3.

Safety factors can be used at this stage of the risk management process related to individual events. The basic idea is, that the analyst links every safety event with the safety factors which failed and safety factors which saved the day. Especially concerning the positive experiences, it is usually best to concentrate on the key safety factors which really made the difference. This way, one gradually learns more and more about what kind of safety factors fail in particular situations, or related to specific ship types or conditions, etc. One also learns what are the positive safety factors which provide the necessary safety margin in different situations. Due to their nature, the safety factors point quite nicely to potential improvements in the system: for example, if "communication between the ship and the outside world" or "knowledge" get high on the ranking list, is not difficult to start planning corrective actions. The list of safety factors is also quite manageable as there are not so many of them. Because the safety factors are quite different from typical taxonomies they also provide new perspectives to existing safety challenges.

The safety factors can be particularly powerful when used in combination with the event risk values. The event risk allocated to a specific event can be allocated to all the safety factors that were linked to that event - either positively or negatively. This provides a powerful way to assess the importance of different safety factors - not only through counting events but through the actual risk at stake. This is illustrated in Figure 19. The horizontal axis lists some railway safety factors through their code names (e.g. 28 = situation awareness; 24 = application of procedures and knowledge). The blue bars (with the scale on the left) indicate in how many events the SF's failed and the thus cumulated event risk is indicated by the checkerboard bars (with the scale on the right). Note the significant difference between these two rankings: the difference comes from the fact that the event count is only sensitive to the *number* of events while the (cumulated) event risk is also sensitive to the *potential severity*. The latter

can support risk-informed decisions while the former cannot. The sample consists of 533 events and only top safety factors by event count are presented in the figure.

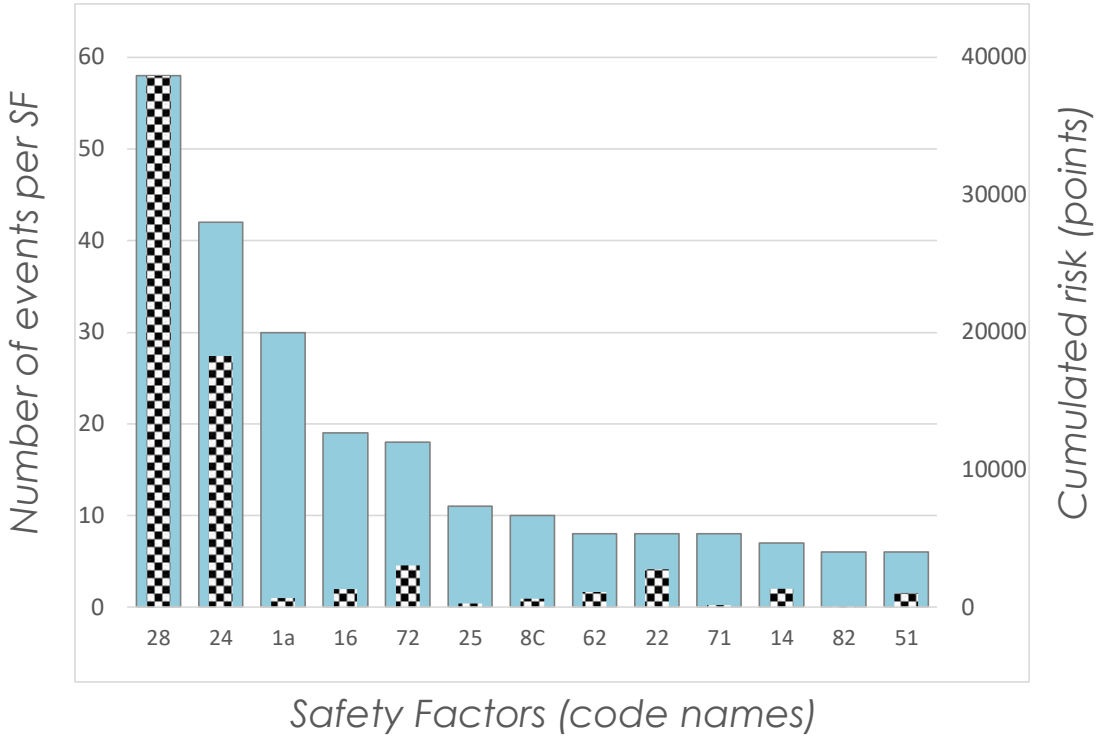


Figure 19. Example of ranking some Safety Factors by event count (blue bars) and cumulated event risk (checkerboard).

Safety factors can be used in multiple ways at the later stages in the risk management process. For example, they can provide a way to make sense of specific safety threats or scenarios through the identified safety factors in the events related to those threats/scenarios. Safety factors can even be used as a base for operational observation: e.g. thanks to their formatting as positive statements they can be used as guidance for where to look for operational resilience. In summary, the safety factors support the requirement (RA-5) by highlighting which factors have "saved the day". They also point to areas where more resilience can be built in the system (requirement RT-2) and help focus resources more precisely on the key topics (requirement C-2).

The safety factors should not be used as another reductionism-driven taxonomy or recipe for success. In reality, there are many non-linear interactions between the different safety factors and dealing with these factors one by one is a simplification of the complex reality. For example, severe fatigue ("vigilance level") may quickly wipe out all the benefits of the fantastic teamwork and communication skills that the crew has on a normal day. Similarly, the impact of three safety factors acting simultaneously may be drastically higher than the sum of three factors taken separately.

Despite the aim to stay at functional level, one can see that some safety factors are at a more detailed level than some others. The list of safety factors could be created in multiple ways and it is difficult to claim that one set would be more justified than another. It should also be noted that the whole discussion here is focused on safety factors at the operational level. If one wanted to create safety factors for the organizational level, that would be possible too but would imply another set of safety factors. Finally, the term "safety factor" may cause difficulties as the same (and similar) terms are used in multiple meanings. The term "safety function" would not have been fully justified as not all of the factors are functions.

3 Risk analysis

3.1 Data Integration and building knowledge

The objective of the data integration step is to make sure that all relevant information contributing to risk assessment and decision making is gathered and that all different types of information are combined, enabling the shared understanding of risks and the creation of the risk picture. In other words, this step focuses on enhancing the Strength of Knowledge (SoK) and thereby addresses the requirement RA-2.

On one hand, there needs to be focus on the *breadth* of knowledge: getting all the relevant information together; and on the other hand, there needs to be focus on the *depth* of knowledge: making sure that the available information is used in the best possible way. The former concerns not only data but also access to the right people who may not be part of the organization carrying out the risk assessment. The latter implies constructive discussions between experts and being able to bring in as much of the tacit knowledge as possible, without falling in the traps like groupthink.

The risk identification step has produced raw safety data and preliminary analysis results based on this data. These results can be in the form of various types of studies providing insights to the different aspects of operational safety and risks. Another type of result is the identification of safety issues, or at a more generic level, threats. Thanks to the event risk assessment and the safety factors, both the role even data and the preliminary analyses based on them can be enriched through the use of these two methods. Building on the list of sources in Chapter 7.2, there are now at least six broad categories of relevant safety information to consider:

- Safety event data, benefiting from safety event risk assessment and the safety factors.
- Preliminary studies and analyses, potentially benefiting from event risk assessment and safety factors. Some of these studies could come directly from operators based on their own data analysis.
- Safety issues. These are identified safety concerns which may be risk assessed using the safety issue risk assessment, covered later in Chapter 7.3.3. Compared to safety events, safety issues are at a higher level of information as they typically combine specific hazards with specific conditions and contexts.
- Oversight information. This covers information flowing in from audits and other oversight activities. It is good to keep in mind that the people who carry out the oversight activities know more than can be written down in reports, such as audit reports. Therefore, having access to these people may be at least as important as having access to the relevant documents.
- Information, knowledge and experience brought in by experts and through existing studies and research findings. The experts bring in both their operational experience from a certain domain and their specific knowledge about safety in the domain. Additionally, a lot of research has been carried out in the domain of transport safety and such results may be useful depending on the topic being discussed.
- Understanding about how things can evolve in the future. Ideas about what the future could look like are presented here as a specific point because all the previous safety information categories reflect historical information. Yet, risk assessments and any actions taken always target the future. Consequently, it is good to be reminded of the fact that the future may be different from today and despite the difficulty in trying to guess what it could look like, one should take a conscious effort to integrate thoughts about future trends in the risk management process. Here one has to consider not only safety related issues but also general changes in the environment which may gradually have indirect influences on the transport system and its risks. For example, the availability of powerful lasers and more recently inexpensive drones are having an increasing impact on the safety of commercial aviation.

Based on this list, it can be seen that the designed process can digest virtually any type of safety data/information as an input. The proposed practical way of integrating all the data together is through

so-called *risk workshops* which will be discussed below in Chapter 7 .6.1. In fact, most of the remaining steps in the process can be carried out in the risk workshops.

When discussing a specific topic in a risk workshop, all the relevant information is reviewed and discussed trying to ensure an adequate level of breadth and depth. The thoughts and conclusions can be recorded for further use. The result should be an enhanced understanding of the topics under study and a transformation from data to information. At this point, the focus is solely on understanding the situation and any discussions about solutions should be postponed to a later stage, so that the analyst group does not fall in the trap of making fast conclusions based on a quick analysis, possibly using only a part of the available information.

As a part of the work on "future risks", it is good to dedicate time on potential surprises. In addition to brainstorming what kind of surprises could be experienced, one innovative way is to use the safety factors as the starting point. For example, the team could take the safety factor "*availability of timely and reliable information, between the ship and the external world*" (see appendix 3) as the starting point. An event, where this safety factor would fail, would indeed be a surprise. Such a surprise could now be examined through the rest of the risk management process. The advantage of using the safety factors is not only that a number of potential surprises can be identified, but also that the starting point is a loss of a key function (especially in the case of the fundamental safety factors) and this can be obtained without working on the hundreds of scenarios which could lead to this functional loss, and without entering into arguments on whether such a loss is "realistic" or "possible".

The data integration step has several deliverables. During this step, many of the threats, scenarios and safety issues can be identified and named, even if this is also possible in the later steps. All the material available about these elements can be reviewed and a common understanding established and recorded. This provides the content for the risk picture discussed under the next section.

3.2 The Risk Picture

The data integration step deals with several distinct types of information and also different kinds of issues: some can be large overall issues in the transport system and in the other extreme, some are just single events that have occurred. There has to be a way to make sense of all this information and organize it in a way which supports decision making. The proposed way of doing this is to construct a so-called risk picture. The risk picture is a two-dimensional presentation illustrating threats, scenarios and individual events visually. The risk picture aims to provide a holistic overall picture of the risk situation. It is also the starting point for the decision-making process. It helps analysts see the *nature* of different threats - for example, it can highlight which threats belong in the high-impact low-probability area and thus require a special approach as discussed in the context of black swans. The risk picture also reflects the criticality and strength of knowledge of different threats.

In order to support decision-making, the risk picture needs to *highlight differences* between different threats and scenarios. It also needs to address the difficult matter of *acceptability of risks*. As discussed, the acceptability is not a simple correlation with risk - it is a very relative notion depending on the context: different things are more or less acceptable, sometimes even in a seemingly irrational way. Moreover, as stated before, it can also be considered that a particular risk treatment plan is accepted, rather than a particular risk. From the risk picture point of view this means that the acceptability paradox cannot be solved by simply drawing a line through the risk space.

In the context of different modes of transport, each mode of transport could have their own risk picture with their own threats. However, one can also bring together the four modes of transport in a single risk picture. This offers the possibility to try to optimize risk management at the level of the whole transport system which should be the ultimate objective. This implies the compatibility of the local risk pictures.

Going back to the modern definition of risk, it is clear that presenting risk simply as a number would be very simplistic and provide a very poor level of knowledge. Different combinations of probability and

severity may produce the same risk level, mathematically speaking, but to be very different types of risks to manage, and therefore could deserve different priorities. An already-cited example would be the high-impact low-probability threats who may have to be given a high priority despite their apparently low risk value as a number. One wants to understand the nature of the risk. Therefore, for the risk picture, a *vector* presentation of risks is adopted. Risk is presented in two dimensions through the two vectors of severity and probability. The actual risk value would be the product of these two vectors and create a third dimension which is not presented. However, the location of the threat within the two-dimensional space fully determines the risk. Thanks to this two-dimensional presentation is also easy to present the notion of strength of knowledge.

The two axes of the risk picture are the same as used for event risk assessment: severity vertically and probability horizontally. The severity scale is the same point scale as described for event risk. Similarly, the probability value is again naturally a number between zero and 1. Theoretically, the probability scale continues to infinitely small values in its lower end but stops to the absolute value of 1 on its higher-end. Therefore, it makes sense to make the vertical axis intersect with the horizontal axis at the probability value 1. The range for both axes is so large that one is practically obliged to use logarithmic scales. Obviously, the risk picture can be arranged graphically in different ways while respecting the same overall principles.

Now that the two axes are defined, the risk picture can be filled with the desired content – the objects placed in the risk picture can be called (risk) *elements*. The largest elements to be placed in the risk picture are the threats. The threat would then typically contain several scenarios. There is usually a fair amount of uncertainty about severity values and the probability values related to the threats and the scenarios - despite the fairly harsh "precision" of the logarithmic scale. Therefore, for any threat, instead of selecting a representative value of severity and probability, it is proposed that a *range* is used in both cases. This means that the threat is symbolized not by a point but by a rectangle. The estimated severity range gives the height and the estimated probability range gives the width of the rectangle. Another possibility would be to use crosses instead of rectangles.

Another symbol or color can be used for the scenarios. Logically, the scenarios should be located within the boundaries of the related threat – i.e. the rectangle. The convention applied here symbolizes scenarios with a hexagon. Threats could overlap, and a scenario could be related to more than one threat. One must be humble in acknowledging the uncertainty and imprecision in estimating the numerical values related to the different elements. The amount of time and resource to obtain more reliable estimates needs to be balanced with the actual benefit. It may be enough to get a rough idea of where the different threats are located in relation to each other. It must also be noted that in many cases the estimation becomes almost impossible when one reaches the area of very low probabilities. An exception could be failure cases related to detailed system safety analyses which could provide q probability estimates also in the region of 10^{-9} to 10^{-12}.

In addition to highlighting the uncertainty by the use of ranges (and therefore rectangles), the strength of knowledge could also be visualized in another way, for example through the line width of the rectangle or (line or fill) color of the rectangle.

Placing all known threats and scenarios in the risk picture gives the overall picture of risks. Even very different types of scenarios from different modes of transport can be placed in the two-dimensional picture because the used scales are universal. As there are no absolute limits for what is acceptable in terms of risk, it is very helpful to be able to compare the locations and uncertainties of different threats within the same type of activity - therefore subject to similar tolerance levels to risk. For example, it makes sense to compare different threats and scenarios related to commercial aviation between themselves; and threats and scenarios related to leisure marine navigation between themselves. The acceptability levels of risk would be quite different from one domain to the other but could be assumed to be very similar within a single domain.

There is yet one more important element to be added in the risk picture. Interestingly, as the event risk assessment shares exactly the same two-dimensional risk framework, all the individual safety events can be plotted in the risk picture using their event risk assessment values. As they reflect events that occurred in the real world, they can bring in a valuable reality check compared to the threats and scenarios placed in the risk picture. It should be noted that the concept of probability is slightly different for events compared to threats and scenarios. For threats and scenarios, the probability refers to the future: what is the probability that this threat *materializes in the future*. For historical events, the probability refers to the probability that the event *would have escalated* into an accident outcome. However, these two concepts work perfectly together within the same risk picture and complement each other. In a way, the threats and scenarios indicate the assumed values while the events illustrate how close to an accident the real-life events have escalated in the past. One can imagine the everyday "as-imagined" operation taking place within the fenced area of the threat-rectangle, and sometimes some events escape from the rectangle and escalate towards the accident, i.e. the vertical axis which also symbolizes the probability value 1 and thus an accident. Indeed, if the data contains events which had actual outcomes, those events would be placed on the vertical axis – and for such events the risk picture is rather a riss picture.

By processing the individual events through the event risk assessment and then bringing them into the same risk picture with threats and scenarios, not only have events been transformed into risk information, but they've also been brought into the overall risk management process in a quantifiable and visual way.

The concept of a Safety Issue is somewhere between a threat and a scenario. A safety issue would be treated exactly like a threat, the only difference is that a safety issue usually has a more specific and limited scope than a threat. For example, landing on short or narrow runways could be a generic threat. An example of a safety issue could be: "night operations to airport X". The safety issue would combine several *hazards* like nighttime (no light, fatigue) and the difficulties of a particular airport (which could include a short runway) but the whole safety issue is specific to that one airport. It could be that this is the most difficult airport in the route network of the airline carrying out the risk assessment. And in practice the safety issue would also be specific to the aircraft type that the airline uses for that specific route. Therefore, safety issues are usually very well defined which helps carry out risk assessments with a little bit less uncertainty (see Chapter 6.3 and Appendix 2). Despite the specificity of a safety issue, conceptually it can be treated similarly to a threat and safety issues can be placed in the risk picture in the same way as threats. By doing so, the safety issues again can be compared with other threats and the notion of acceptability can be tested in relation to other elements in the risk picture. Similarly to threats, a safety issue may have several scenarios linked to it. Safety Issue Risk Assessment (SIRA) is discussed below in Chapter 7.3.3.

In practice, there will always be inputs to the risk management process which come from outside the data streams, as reflected in requirement (C-3). Most typically, for a national safety agency, the political system including the ministry of transport will take initiative and come up with proposals on different safety actions and activities. Despite the political weight behind such proposals, sometimes the real merit in terms of risk and safety improvement could be questioned. Once the risk picture has been established, it can be used to host such proposals as safety issues. This way, one can see where the issue is located compared to other similar safety issues and threats. This can be a better base for discussions and argumentation than treating the proposal alone in a vacuum.

Similarly, even if organizations (like transport operators) are addressed specifically at the level of organizations, in line with requirement (RT-10) (see Chapter 7.5.6), a problem related to a specific operator could also be placed within the risk picture as a safety issue. All this illustrates the flexibility of the risk picture in creating a holistic view to the different threats in the operating field. An example of a risk picture is presented in Figure 20. Threats (T1-T3) are presented as rectangles. The thickness of the outline symbolizes the strength of knowledge related to the particular threat. Scenarios are presented as hexagons. Events are presented as dots and linked with the relevant threats with a line. Note that a

single event may link with several threats. With the chosen direction of the axes, risk is increasing from bottom right to top left, which is also indicated with a colour scale.

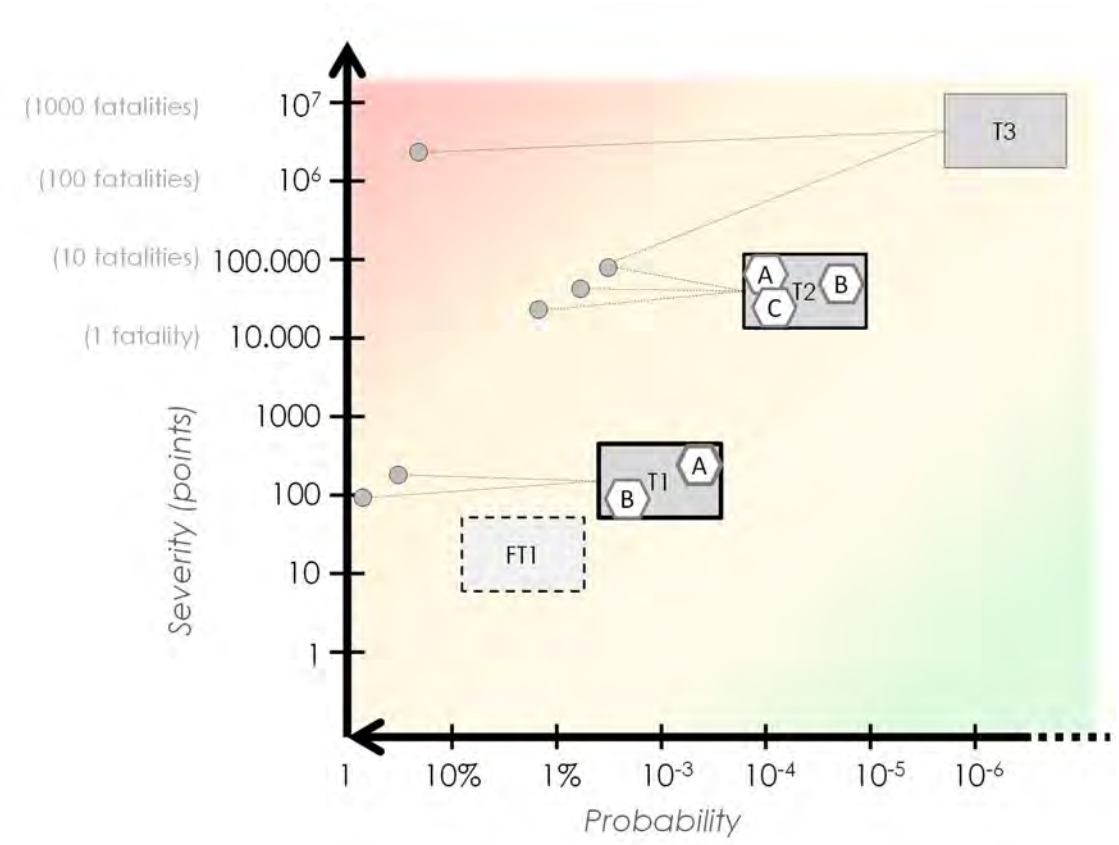


Figure 20. Example of a Risk Picture.

If there are risk elements which are seen to emerge in the *future*, there is a possibility of including them in the risk picture and potentially using specific symbology, like the dotted-line-rectangle (FT1) in Figure 20.

Two more aspects need to be covered before the risk picture becomes a suitable base for decision-making: the different exposure levels to different threats and the balance between risk criticality and the cost of related improvement. The former issue addresses the fact that the risk picture constructed as explained above is blind to the different exposure levels of different threats. For example, threat A may be related to flying a specific model of a helicopter and threat B may be related to a continuous commercial airline operation with a fleet of over 50 aircraft flying almost around the clock. Even if coincidentally the probability levels of the two threats would be identical, in reality there is a much higher exposure to threat B because the helicopter flights using a specific model are comparatively speaking rare. This means that another version of the risk picture needs to be created which takes into account the different exposure levels. This matter is discussed below under Chapter 7.3.4. The latter issue highlights the fact that decisions on action should not be based only on the level of risk allocated on different threats. Some problems are easier to fix than others and sometimes fixing easier problems quickly may produce a bigger total safety impact than fixing the number one threat in terms of risk level, if the solution is very difficult and time-taking. This means that the risk level and the cost of improvement need to be considered in parallel. Here *cost* should be understood in a very wide sense including time, financial cost, resources needed, skill and experience levels needed and so on. This matter will be discussed in the context of decision-making in Chapter 7.4.4. Moreover, interventions should not be considered one-by-one but as a whole, in line with requirement (RT-6).

As a result of the described design of the two-dimensional risk picture, including threats, safety issues, scenarios, safety events and potentially other elements - and using some conventions on the symbology, there is now a holistic picture of all known threats in the operation, addressing the requirement (RA-6). It allows to compare different threats in terms of risk and understand the nature of the threats depending on their location within the picture. Together with the second version of this picture which takes into account the exposure levels, these risk pictures form a useful tool for decision-making. In terms of bandwidth for presenting even more information, there is still room for using other aspects of symbology (e.g. colours, shapes, line styles and widths) to include more information. The design of the risk picture is naturally highly customizable and what has been presented above is only one way of creating the picture. Some choices will become obvious only in the real use of the risk picture with real data - for example, if there is a very high level of data it may require different solutions to avoid extensive cluttering of the risk picture. Implementing the risk picture as a software gives new possibilities to manage different aspects of the picture including the clutter. For example, only desired elements could be presented at any given time. Implementation of the described risk management framework with the help of a software is outlined in Chapter 7.8. From the terminology point of view, the risk picture could also be called a "picture of threats" or "threat picture", so these terms may be used interchangeably.

3.3 Safety Issue Risk Assessment

The original Safety Issue Risk Assessment (SIRA) as created by the ARMS working group was presented in Chapter 6.3. It is argued, that the concept of safety issue and the ability to carry out safety issue risk assessments are important for effective risk management. Typical safety problems have a scope and specificity which matches well with the concept of safety issue, more than with a very large concept of threat (e.g. strong winds) or a single scenario.

However, the original SIRA method has some limitations. First of all, the outcome is a simple judgment on whether the risk is acceptable or not. The accident probability is calculated based on the exposure and the estimated effectiveness of the barriers, the related severity is picked from a few categories, and the combination of these two is compared with predefined acceptable combinations of risk. If the example Excel version of SIRA is used, the answer shows how far from the limit of acceptable the current value of risk is, using five different categories. This judgment is based on the predefined risk limits which is probably one of the key limitations of the method, as literature does not really support a single acceptability limit. This is also reflected in the requirement (RE-1). Moreover, even if acceptability limits were established, the original SIRA method would not facilitate a comparison of different safety issues in the way that the two-dimensional risk picture does.

Another limitation is that in the current example application the severities are entered using predefined categories. Compared to the system of points where any level of fatalities or injuries can be addressed quite precisely, the category system is less precise and particularly weak at the top end of the scale, as 3000 fatalities would be considered the same as 50. Taking into account other severity dimensions like environmental and material damage would not be possible. Furthermore, the original SIRA is not compatible with the scales of the original ERC nor the current updated event risk assessment.

These limitations introduce the need to redesign the safety issue risk assessment. The existing event risk assessment framework and the two-dimensional risk picture give a good base for this. Indeed, both axes of the two-dimensional risk picture can be used as such for the new SIRA. This means, that once the severity and probability values for the safety issue have been estimated, the safety issue is simply placed within the two-dimensional risk picture using the same rules that already exist for this picture – i.e. it is not a point but an area as both values are estimates and become ranges on the two scales. Finding out the location of the safety issue in the risk picture replaces the simple judgment on acceptability. The location of the safety issue can now be used in risk evaluation, both in an absolute sense and in relation to other risks.

The rest of the original SIRA method remains valid. Defining the safety issue first carefully before starting any assessments is fundamentally important. This implies scoping the safety issue carefully so

that the assessment can be more factual. The existing Excel example application can be very useful in carrying out these initial steps of the assessment. It is also possible to use the model from the original SIRA in trying to assess the effectiveness of the barrier system and thereby arriving step-by-step to the probability estimate. Sometimes such a split may be too artificial or sometimes the uncertainty will not allow such a detailed analysis. In such cases a simpler, rough estimate may have to be used.

As discussed above, safety issues treated this way are now compatible with the risk picture and can simply be placed in it among the other threats and scenarios. This way, the safety issues become integrated in the overall risk management process, addressing the requirement (RI-3). If there is a need to examine specific safety issues compared to some predefined limits of acceptable risk, it can still be done either numerically or within the risk picture but it is argued that most of the time the visual answer provided by the risk picture will probably be the most useful outcome of the assessment; not only for allowing comparison visually with other risks, but also because it is able to convey more about the uncertainty involved using the ranges and specific symbology to reflect the strength of knowledge.

Many organizations must establish Safety Cases for specific activities and submit such analyses to various authorities. This may be a requirement imposed by the authorities and/or a mandatory process within the safety management system of the organization itself. An example could be flying to a destination with known military activity around the destination airport. The risk assessments of this kind often struggle to show why the end outcome of the assessment is acceptable, and the author hasn't seen specific *conditions* set for the acceptability, for example having to update the assessment every day. It is argued that placing the risk among other risks will be helpful in justifying a specific outcome of the assessment.

3.4 Exposure rates and the Type II Risk Picture

As mentioned above, a risk picture based on probabilities does not take into account the fact that different threats typically have different exposures. Therefore, another version of the risk picture needs to be created, taking into account the exposures. The severity dimension remains unchanged, so visually speaking the threats, safety issues and scenarios will move left and right depending on their exposure rates.

There is a simple way of creating the exposure-sensitive risk picture: the estimated probability values (for example 1 out of 10,000 operations for the accident to materialize) are multiplied by the exposure to the operation in question, in a certain time period. Most typically, the time period could be one year. For example, if the probability is 1 out of 10,000 and the yearly exposure is 1 million operations, the result would be 100. This result indicates what would *be the theoretical average (or most likely) number of accidents taking place in a year due to the threat/issue/scenario in question*. In reality, the real number of accidents related to the threat could be very different and it would take many years with a stable probability and exposure rate to show that the average number of accidents is indeed this calculated value, for example 100 accidents per year. It must be acknowledged that the actual numerical values are theoretical and often based on rough estimates. However, this approach is effective in creating a risk picture which takes both the probability and the exposure into account and therefore creates a two-dimensional presentation (with severity on the vertical axis) which can be used for prioritizing different elements using meaningful criteria.

The horizontal axis could now be named "theoretical average number of accidents per year" and this naturally refers to each element separately. This version of the risk picture could be named "Type II" risk picture (where the original version would be "Type I"). It should be noted, that there is an important difference to FN-curves. This is because the latter refer to the probability of "one or more" accidents occurring in the given time period. For prioritization purposes the chosen simple method for the type II risk picture is sufficient and easier to calculate.

There are two types of elements which need to be considered separately: elements with very low probabilities, especially ones with high severity values, and safety events. As very small probabilities

are practically impossible to estimate, elements with very low probability values are associated with rough estimates of their probability range. Therefore, also their locations in the type II risk picture are vague. This matter is important especially for elements in the top right corner - in the so-called black swan corner. This is because these elements are high-impact risks and making a too optimistic assessment on their location on the horizontal axis could lead to a dangerously optimistic assessment related to these risks in the operation. As a consequence, a special way to present these elements in the type II picture is adopted. The presentation format of safety events is discussed below and the discussion concerning the elements in the black swan corner can be found in the dedicated Chapter 7.4.3.

As discussed above, bringing the individual safety events into the risk picture gives an interesting perspective to what has actually happened in terms of the operational events. The type I picture is directly supported by the event risk assessment, as its result positions the event on the picture. For the type II picture, the severity values pose no problems as they will remain identical to the type I presentation. However, the probability values for safety events in the type II picture are no longer straightforward.

There is a fundamentally important piece of theory related to event risk that makes events different from the other elements on the risk picture. When done correctly, threats, scenarios and safety issues are defined, so their scope is precise. This means that their exposure is also definable, at least in theory. For example, for a safety issue, defining the time period under study, the vehicle/aircraft/ship types concerned, routes in question and so on, means that it is possible to calculate how many times a year such an operation takes place. In this respect, safety events are different. Their scope is not defined. The fact that a specific type of ship was involved in the event does not tell if the reference scope of ships includes all ship types, only some of them, or only the specific ship type in question. If the event took place in daytime, in January, it does not define whether the reference scope should be only daytime operations or 24-hour operations, nor does it define whether the scope is anytime of the year, wintertime only, or January operations only. The fact that a single safety event may link with several threats, scenarios and safety issues illustrates the same point. The consequence for the type II picture is that it is impossible to find a reference exposure rate for an event.

Another important aspect of safety events to keep in mind is that every event is unique. Therefore, in a risk picture, they should not be combined through a classification into different categories. Safety events with actual outcomes are slightly easier to deal with as the specified consequence with the given severity level actually took place. Events with only potential outcomes may be even more challenging as there is only an estimate of probability that the event would have escalated to the given outcome.

From this theoretical background, one can derive a possible presentation format of safety events in the type II picture. As the exposure dimension is not really relevant for events, their position on the horizontal axis is to some extent a question of convention. The proposed practice is to place them at the value 1, symbolizing that one event like this took place. This means all safety events would be placed on the same vertical line. The probability value of the event risk assessment would no longer make a difference between different events. One could ask, why isn't an event with a 50% probability in a different place than an event with a 100% probability (to have produced actual outcomes). The answer is simply that the exposure rates would move both events left or right, but the exposure rates are not known. So, every event is at the "once a year" value to show that the operational experience includes such an event. It is probably recommendable to make a distinction between events with actual outcomes and events with only potential outcomes. This could be done through symbology, for example using different colors. From the purely pragmatic point of view, presenting a lot of events in the type II picture will be challenging as they will all tend to be in the same area. Yet, every event should be presented individually. This may require spreading out events landing on the exactly same value so that they become visible separately. Of course, there is also the option of presenting events only in the type I picture. In any case, the events are presented in the risk pictures only to give another perspective to the threats, scenarios and safety issues. Events are history. Nothing can be done about them anymore. All possible actions can only address the future and the future is reflected by the other elements in the picture, not the events.

Figure 21 presents an example of a type II risk picture. In terms of threats, safety issues and scenarios which are not in the very low probability area, calculating their new positions in the type II risk picture is straightforward. Threats and safety issues have moved horizontally based on the exposure rates (thick dotted lines indicate displacement). A few safety events with actual (red) and potential (white) outcomes have been plotted. If yearly exposure is used, then the resulting position on the horizontal axis indicates a theoretical number of accidents occurring in a year related to the threat in question. In this example, T2 has high exposure (e.g. related to every passenger train operation) and T1 has a lower exposure (e.g. related to hobby flights with autogiros) as well as the safety issue (they are often quite specific, and thus easily have lower exposures).

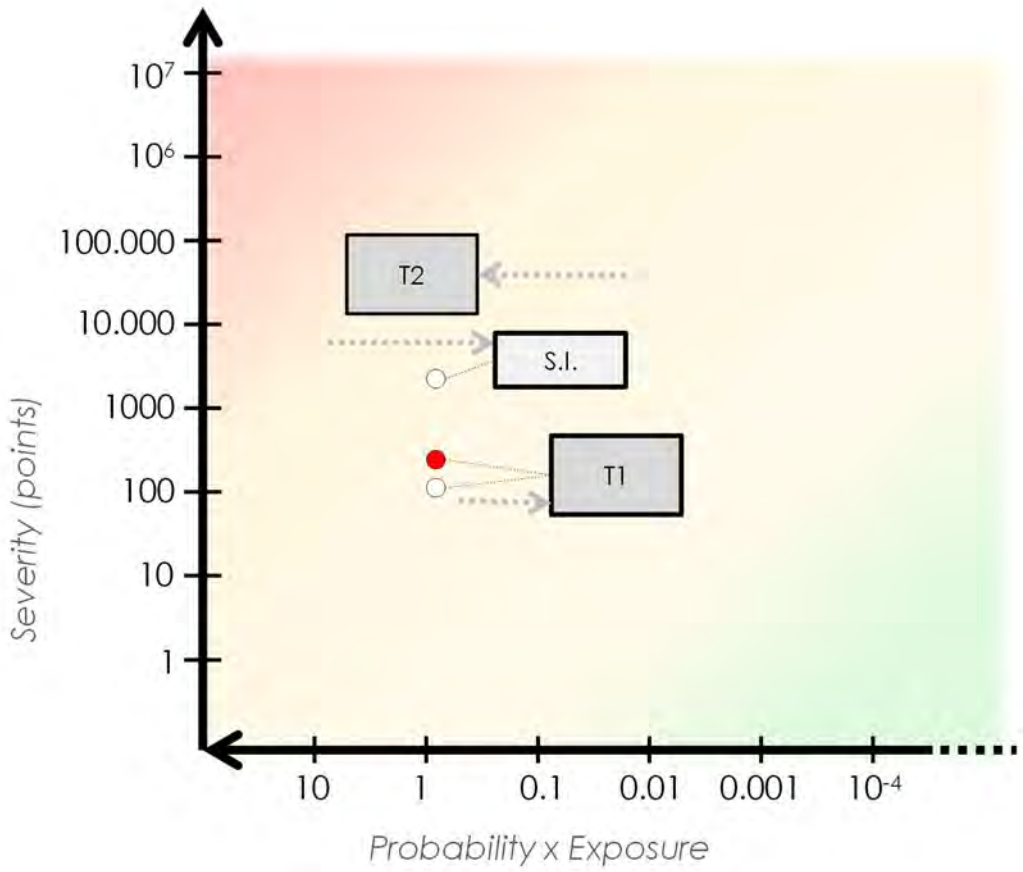


Figure 21. Type II Risk Picture.

So far, the implicit assumption has been that all elements can naturally first be placed in the type I picture and then transferred to the type II picture. In real life, it may be that some elements are easier to place directly in the type II picture. In some cases, the probability information needed for the type I picture may be difficult to assess or not available at all. For example, in the context of road safety, the yearly figures for different types of accidents are typically available whereas the specific probabilities of individual scenarios may be very difficult to obtain.

With the type II risk picture, the different exposure rates of different risk elements are now taken into account. Consequently, the different elements are now comparable both in terms of severity and probability - which has now been transformed into an average yearly accident frequency. This new vector presentation can still be interpreted as risk with the probabilities now having been weighted by their exposures. The risk picture provides an aid for decision-making. From a basic risk management point of view, risk is starting at its lowest value in the bottom right corner and increasing to its maximum value in the top left corner. A simple prioritization based on risk would follow this logic. In the proposed approach, however, there are further aspects to consider in the decision-making.

One could criticize the risk picture by saying that it reflects the old risk perspective, referring only to severity and probability. However, it must be remembered that the uncertainty is reflected in several ways in the picture and that the full rationales for each element in the picture have been developed during the knowledge building stage, and this information is recorded and is therefore available, including all the key assumptions taken. Finally, to address the new risk perspective fully, the topic of black swans is also addressed (see Chapter 7.4.3).

4 Risk evaluation

This chapter covers the use of the risk picture for decision-making. The ways in which different risks within the picture can be interpreted based on their location is discussed first. The second section covers other aspect of risk analysis based on the risk picture. Section three explains the special treatment of low probability-high severity threats. The fifth section sums up the process of decision making and acceptability of risks. As will be seen, the risk picture itself is a good basis for understanding the current situation in terms of risks. However, final decisions about actions are not recommendable based on the risk picture alone. This is because action should not be driven solely based on risk criticality but also guided by how easy or difficult it is to achieve the desired change in the surrounding transport system. As reflected in requirements (RE-1) and (RT-1), risk evaluation, decision-making and risk treatment need to overlap to form a single holistic decision making process. Consequently, Chapter 7.5 needs to revisit the decision-making topic.

4.1 Understanding the different areas in the risk picture

The description about the different areas of the risk picture here assumes that the direction of the axis is as presented in Figure 22. This places the corner with the highest risk levels to the top left and the so-called black swan corner to the top right. As mentioned above, the risk picture could be presented in different ways; another way would be to make the horizontal axis run from small values in the left towards the bigger values to the right, and consequently place the vertical axis to the right side (still at value 1 of probability).

The risk picture is naturally a continuum both in the vertical and horizontal sense. However, the idea here is to give a sense of what's the different areas in the risk picture mean in terms of the nature of the risks. There are no strict lines between these areas, which is symbolized in Figure 22 by the thick gray borderlines separating the different areas.

The two most obvious areas to describe are the top left and the bottom right corners. These are the extremes in terms of risk. The top left corner is the area where one does not want to see any operational threats appearing: this area is characterized by high severity and high probability. The bottom right area is the area which could be ignored if the perspective is purely from the risk management point of view: resources should not be wasted on low-impact low-probability threats. When the additional dimension of "resources required for improvement" is taken into account, this general principle may not hold as strictly anymore, though, and this aspect is discussed in the next chapter.

Elements in the bottom left corner should be unimportant most of the time. There could be exceptions for very high frequency elements. The difference compared to the bottom right corner is the potentially high probability and therefore potentially high frequency of occurrence.

Moving to the medium severity range, the left side of this area is the high probability - medium severity area. Generally speaking, this is not a desirable area to be in, as the risk will still be very high due to the medium severity levels combined with high probabilities. On the right side, the risk goes down due to the lower probabilities. However, probabilities in this area may be difficult to estimate as they may be getting very low and if the severity levels are relatively high some of these elements may still be important: the top edge of this area touches the so-called black swan corner.

The top right corner is the area for risks with high potential impact but assumably low probability. Some of these risks would be well identified and just happen to belong to this area. However, some of the risks in this corner are the so-called black swans or the unknown unknowns. Due to the special nature of these risks, they need a special treatment. This topic is covered below in its own section.

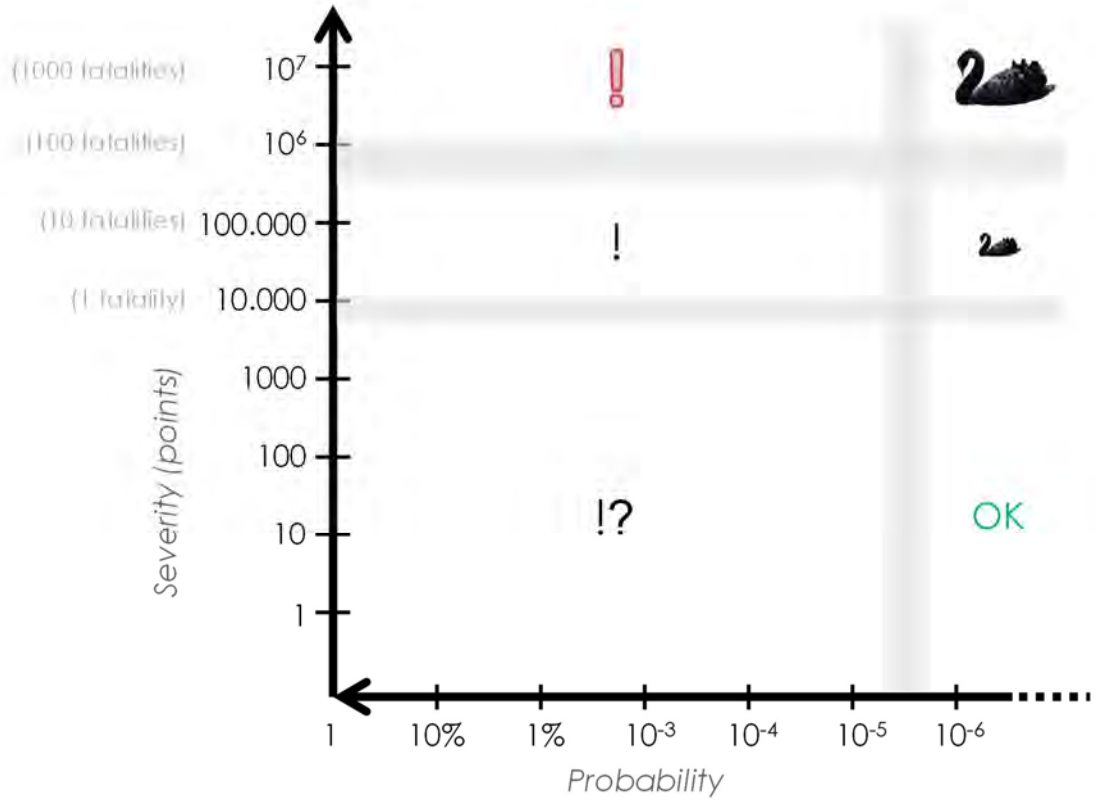


Figure 22. Different areas in the Risk Picture. It is good to note that areas symbolized with a black swan do not contain *only* black swan-type threats.

As mentioned above, splitting this two-dimensional space into the six areas is a simplification but helps to illustrate the different natures of risks in the different areas. The vertical divider between the left and the right side has been placed between 10^{-5} and 10^{-6}. There are several different reasons why this range can be used as the transition area. First of all, it is the limit of ultra-safe systems, i.e. the safest man-made systems have reached the safety level of less than one accident in 1 million operations (Amalberti 2001). It is also the range of probability values where, at the latest, most of us struggle a lot to feel the difference between the consecutive probability values, e.g. 10^{-6}, 10^{-7} and so on. Interestingly, it also happens to be the reliability level which can be reached by technical solutions like automation. According to Amalberti (2001), the system cannot be made safer beyond this point simply by introducing more automation.

There are two dividers horizontally. The lower limit has been set to the value of one lost human life, i.e. 10.000 points. Below this limit the consequences can be considered relatively minor. The other limit has been placed to the region of severity which marks the transition from an accident to a major catastrophe, so between 10 and 100 lives lost – naturally a fully subjective judgment.

In addition to the *location* of the threats, safety issues, scenarios and safety events in this two-dimensional space, it is possible to carry out some further analysis on the elements to gain further insights on these risks. Such analysis is covered in the following section.

4.2 Further analysis of the risk picture

Elements which have been placed in the risk picture have already gone through the integration step when knowledge about these elements has been developed. Once the elements have been placed in the risk picture there is however an opportunity to gain some more understanding on the various elements.

One important source for further information are the safety events. They can be linked to different scenarios which then link the events with different threats and safety issues. Naturally, a particular event could be linked with several scenarios and therefore with several different elements within the risk picture. Now, understanding gained from a cluster of safety events could be used to understand the related scenarios, safety events and threats better.

Safety events have normally been categorized using some standardized event categories. One advantage of this is that the categories will probably help link the events with scenarios. Additionally, they may provide other interesting details. For example, it could turn out that all safety events related to a certain scenario have taken place in the nighttime. This observation could be used to understand the scenario better and potentially focus especially on the nighttime scenarios.

Ideally, safety events would also have been linked with the safety factors. Looking at the distribution of safety factors linked to a certain scenario can be helpful in at least two ways. First, the safety factors help gain an understanding on which factors may play key roles within the dynamics of creating events within a given scenario. Secondly, the safety factors may also give useful hints about where the leverages for improving the situation may be found. Importantly, safety factors can produce both positive information on the success factors and negative information on failures.

Another way to gain further information from the risk picture is to focus on the *overlap and interaction of different elements*. For example, how much are safety events shared between different scenarios and scenarios shared between different threats? Are there obvious or suspected interactions between the different elements? Such observations may help in finding courses of action which have an impact on several different threats simultaneously, achieving positive changes rather than unwanted or unintended side effects.

4.3 Risks in the Black Swan corner

The top right corner in the risk picture gets its name from the black swans which belong to this area. As mentioned above, there are also other elements than black swans in this area, namely risks that are more or less well-known but happen to have a high potential severity level combined with an assumed low probability of occurrence. In fact, there will be no black swans of type a) (see Chapter 2.3) placed in the risk picture because by definition they are unknown. At best, one could introduce vague threats such as "something that makes an aircraft lose all engine power for the rest of the flight" without specifying what exactly would produce this effect (and most likely against the opinion of many experts who may argue that this is "impossible"). For type b) risks the situation is similar, even if one could argue that they could *become* known if a good knowledge building process is applied. The name "black swan corner" thus does not mean that this corner would host *only* black swans, but reflects more the notion that *also* the unknown, invisible black swans would be lurking in this area.

In line with requirement (RE-4), due to the potentially very high impact of these threats, one cannot simply dismiss them thanks to their assumed low probability. Therefore, the threats in the black swan corner are given *a special treatment: the general idea is that all risks in this area need to be addressed irrespective of the assumed probabilities*. This obviously goes against the more common mechanistic way to use a risk matrix or a heat map.

Several challenges arise. Perhaps the first practical question is, how to present these elements in the risk pictures, and especially in the type II picture which is based on knowing the probabilities and exposure rates. The aim is to illustrate visually that there is great uncertainty about the value on the horizontal scale. One way to do this is to stretch the content of the top right corner towards the left (see threat T3

in Figure 23). Finding a suitable presentation format may be challenging if there is a lot of content in the top right corner. Often the probability values for these threats even in the type I picture could be close to an educated guess.

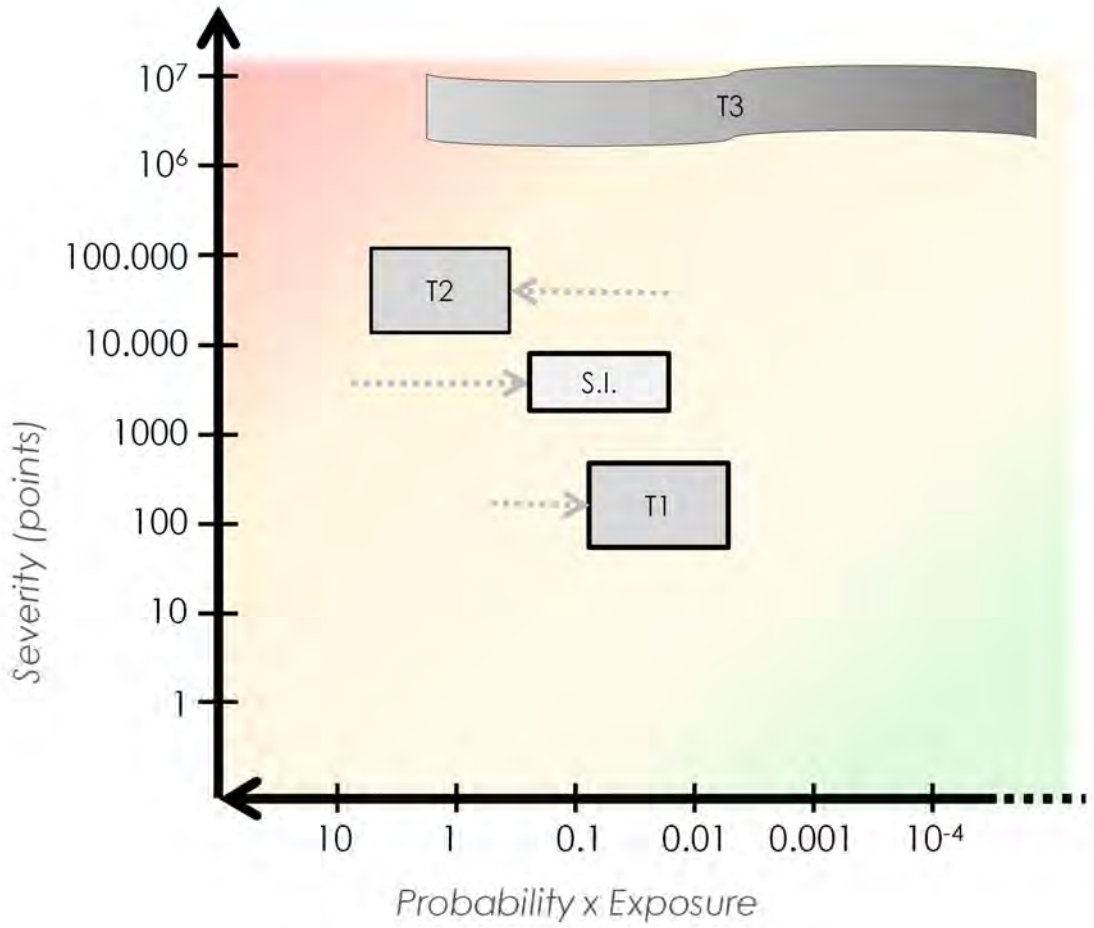


Figure 23. Presentation of threats in the black swan corner within the Type II Risk Picture.

A second challenge is to define a stop rule: which threats should be included in the risk picture and which should not, knowing that in such a low-probability zone there could be a large number of candidate threats. A generic rule could be to include key operational threats which are raised in the risk identification stage. Typically, this would leave out purely design-related threats, such as structural failure due to exceptional external loads.

In the context of the risk picture, the most important point about these risks is to identify them belonging in this distinct group of threats so that they will be dealt with separately during the decision-making and risk treatment stages. Treating these risks is covered in Chapter 7.5.5.

4.4 Acceptability of risks and Decision Making

It is good to remember that at the end of the day risk assessment should contribute to better decision-making. In terms of risks one has to decide if intervention is necessary and if the answer is yes, then what kind of intervention should be used. The former question could be associated with the acceptability of risks. The latter has to do with different risk treatment options. As will be seen, in practice these two questions can be deliberately merged by creating alternatives and then choosing the alternatives that provide the best combination of acceptability and risk treatment.

If a mechanistic view to risk acceptability would be adopted, it would be possible to draw lines in the risk picture, indicating the limits of acceptable risk. If the risk picture is constructed with both severity and probability scales being logarithmic and symmetric, a line of constant risk would go diagonally from bottom left to top right. It should be clear from Part I, however, that this is not the adopted approach. The established requirements for risk evaluation are:

- (RE-1) The decision process should be such, that the decision on the acceptability/priority of a risk or a solution can be done in a *holistic manner*, *taking into account the assumptions* and the *strength of knowledge*, while paying attention both on the risk and on the *costs of its treatment* and considering the *various priorities, not only safety*.
- (RE-2) Consequently, ideally the risk picture should be able to host non-safety risks (e.g. financial, reputational risks) in a compatible manner.
- (RE-3) Decisions on risk acceptability need to be done by humans, due to their capability to consider ethical principles, various conflicting priorities and make value judgments; i.e. the risk management methodology does not need to (and should not) try to produce the final answers directly.
- (RE-4) The *black swans* need to be addressed and the high-impact–low-probability scenarios must not be dismissed solely due to their low probability.
- (RE-5) Ideally, the risk *aversion policy should be left flexible* (i.e. not pre-determined).
- (RE-6) It should be possible to apply specific industry references (e.g. such as by IMO, RSSB or EASA) related to risk acceptance or severity assessment (e.g. "1 fatality corresponds to 10 severe injuries").

It becomes clear that if one wants to optimize the decision-making, risk acceptance cannot be fully separated from risk treatment. In line with Paté-Cornell (2002), there may be clearly unacceptable risks and clearly acceptable risks, but often one is between these extremes. In this case, optimization would mean considering the risks, the alternatives for their treatment and the costs (in a very large sense) of the alternatives simultaneously as the basis for decisions – as Aven et al. (2007) point out: "we do not accept a risk, but *we accept a solution with all its attributes"*. Adaptive risk management takes this approach even further because some of the inputs for decision-making will only be available as a result of the experiments, and the decisions themselves will also be done step-by-step over a longer period of time.

The focus then becomes providing the necessary elements for this new type of decision-making. To address the requirement (RE-1), the risk picture provides a holistic view of the risks and thanks to the knowledge building the assumptions and background information should be available - and ideally not only on records, but in the minds of the people who carry out the evaluation. The estimated costs of addressing the various risks can be presented in the risk picture, as stated in requirement (RT-1).

Figure 24 shows an example of how this could be realized. Arrows show where the Actions (A) and Experiments (E) are expected to push the Threats (T) and Safety Issues (S.I.). Coloured marks indicate specific challenges like long lead time (e.g. A1) or high human resource requirements (E4). A convention can be used for quantities, for example, one blue mark equals roughly six months of lead time. Experiments (in dotted lines) should by definition be short in time and low in cost. Such a picture could of course be constructed in many different ways and a computerized version would probably be the ideal solution.

Risk assessment is now supported by the type I and type II risk pictures and the visualized costs of various alternatives. Different locally customized approaches and rules can be applied to support decision making. For example, existing transport operations which are generally recognized as safe enough within the societal context can be used to anchor various points of currently acceptable risk within the transport system. Naturally, it is also possible to make judgments on individual risks using the probability and severity values, e.g. "are we ready to accept losing 100 people once every five years due to this accident scenario"?

Generally speaking, the type II risk picture is the correct reference for comparing risks and assessing their acceptability, as the exposure needs to be taken into account. On the other hand, it could be argued that a particularly risky part of the operation should not become acceptable only because of its low exposure: every single operation could be expected to be at an acceptable risk level also as an individual operation. In this case, the risks should be in an acceptable area in both type I and type II pictures.

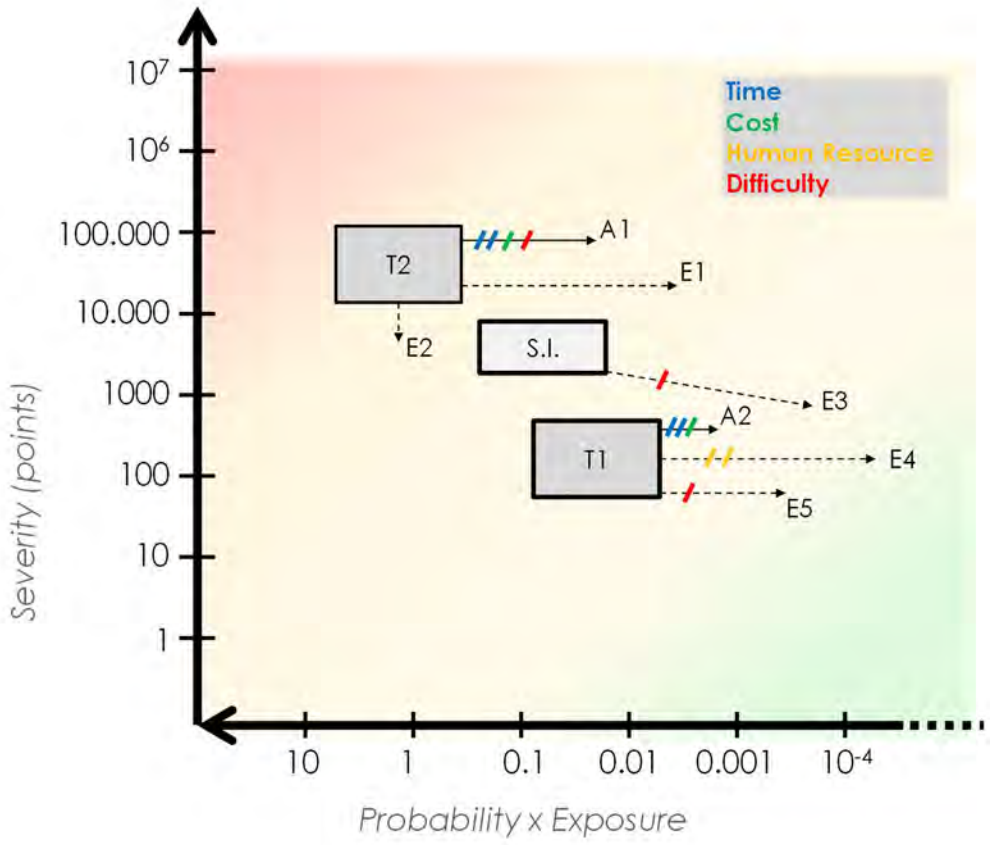


Figure 24. Presentation of potential interventions and their cost dimensions.

Besides the clearly un/acceptable risks, risk analysis based on the risk picture could identify preliminary targets for "desired change". These would be elements seen as "strong candidates" for intervention, in order to move them to a more favorable are in the risk picture. As discussed, the final decisions would be made in an integrated way, taking into account also the estimated intervention costs (and potential for adaptive experimentation).

In the case of a multimodal transport safety agency, each mode of transport can develop its own risk picture, desired objects for intervention and their cost elements. If such an agency wants to optimize its resources across the whole transport system, it needs to bring all the results from the different modes together on a *multimodal risk picture*. This requires involving some key people from each mode of transport who know the backgrounds of the various risk elements.

Whatever the approach, one should also pay attention to the aspect of strength of knowledge, in other words the uncertainty related to the different elements on the risk picture. As discussed in Chapter 2.2, the assumptions behind the probabilistic risk assessment are an integral part of the results, and the SoK reflects the uncertainty associated with these assumptions. The SoK related to various elements within the risk picture should be clearly highlighted (as mentioned in Chapter 7.3.2). In line with Aven (2013a), lower SoK scores should "upgrade" the related risks. Using the risk picture, the SoK can be considered along with other criteria, and taken into account smoothly, as there are no distinct "risk classes" but rather a continuum from low to high risk.

With the multiple dimensions involved (severity, probability, uncertainty, position related to other risks, area within the risk picture) in the decision-making, it is difficult to prescribe a simple algorithm for prioritizing the different elements. At the end of the day, it is a judgment involving several societal values. The recommendation is to do such judgments collectively, bringing as many diverse perspectives as possible, and recording the decisions and their background for future consultation. The practice of running risk workshops is considered well adapted for such type of work.

Another aspect of the decision-making has to do with the interactions between different risks and different actions. There may be positive and negative synergies which will have an impact on what is the optimal set of actions for maximizing the positive safety impact with the chosen resource investment. This topic will be discussed in Chapter 7.5.8 which addresses the Transport System level perspective to decision making.

If a new priority is proposed from outside the main process (e.g. coming from the political system as a knee-jerk reaction to an accident), submitting it to the same scrutiny as other elements in the picture can be quite useful for testing the real criticality of the proposed priority. This may help in modifying the proposal so that it becomes more justified at the level of the whole risk picture.

References are sometimes made to so-called *weak signals*. It is worthwhile noting that a classic risk management approach will systematically ignore weak signals for the very reason that they are weak. If there is a desire to try to capture weak signals, that will have to be done through a distinct process. However, as mentioned in Chapter 4.2.2, it is questionable if the so-called weak signals are ever detectable in an anticipatory mode. In a complex operation, a lot of things happen all the time creating a constant noise (of information). Some of the items within the noise will turn out to be significant in a future safety event but it is very questionable whether such items could have been detected and considered significant, before the emergence of the related event. If the so-called weak signals can only be detected afterwards they are not useful from risk management or safety management point of view.

While it is not claimed that decision-making becomes easy, the point is to illustrate that the risk picture can support the decision-making in a valuable way. The requirement (RE-1) has been addressed, except for the very last part considering other than safety priorities. The idea is that there needs to be a balance between safety priorities and other priorities. This can be achieved in two different ways which can both be used in parallel. First due to the very generic nature of the risk picture, almost any types of risks can be included – also non-safety risks, e.g. financial risks. In this case, such risks will benefit from the same decision-making process as for safety risks, and the requirement (RE-2) also gets fulfilled. The knowledge building size for such risks would have to be carried out among a suitable group of experts, which would certainly not be the same crew asked for safety risks. Naturally, a key requirement is to be able to transform the severity on the point scale, as discussed in Chapter 7.2.1. The second way to address non-safety priorities is to challenge the planned interventions from the perspective of the other priorities. This aspect is discussed below in Chapter 7.6.1.

The requirement (RE-3) is about humans taking the key decisions. It is obvious, that the proposed method aims at supporting decision-making by humans but does not produce conclusions in itself, or even recommendations for action.

The requirement (RE-4) is addressed by presenting risks in the black swan corner in the type II risk picture in a way which highlights their presence irrespective of the original estimated probability values. Risk treatment based on resilience building is also discussed specifically under Chapter 7.5.5. Together with the notion of uncertainty, the specific treatment of surprises comprise the new risk perspective reflected in requirement (RA-1).

There is no in-built rule for risk aversion in the risk evaluation approach, so the decisions are left entirely to the people involved and the effective risk aversion policy could be fully flexible and changed as necessary, which meets the requirement (RE-5).

More generally, the risk picture enhances the premises for good decision-making in several ways. If decisions are made case-by-case without seeing all the different risks simultaneously there is a danger of inconsistency in risk decision-making. Thanks to the risk picture, a number of risks and safety issues can be visualized simultaneously thereby providing references for what are currently accepted risks. This helps in deciding which risks are not acceptable and where these risks should be moved within the risk picture. This also gives an idea of what level of cost the decision-makers might be willing to spend for reducing certain risks. The risk picture also helps in comparing different risk treatment alternatives as these can be visualized within the picture. Different risk treatment options and their combinations can be assessed in terms of benefit and cost, also taking into account potential interactions between these interventions and their potential side effects on other risks**.** All this contributes to a more optimized and more holistic risk decision-making.

Finally, it is very easy to implement various industry practices thanks to the point system on the severity scale. For example, if minor injuries are considered 100 times less severe than fatalities, the former would be valued at 10.000p/100 = 100 points. Similarly, in line with the EASA principle to consider protecting third parties on the ground more than people in the aircraft, this could be implemented in the risk assessment by adjusting the severity points related to these groups of people. For example, it could be decided that all severity figures for third parties on the ground are increased by 10%. Such practices can be implemented both at the level of the event risk assessment and for the main risk management process using the risk picture, addressing the requirement (RE-6).

5 Risk treatment

If the transport system was an ordered linear system and if safety was the only concern, going from the risk picture to the consequent actions would be straightforward and easy. In fact, the target areas for improvement could be identified first, separately from any action, using risk as the sole criterion for establishing a priority order for addressing different risks. Actions would be defined and launched for all the items judged too risky. If resources were taken into account, the priority list would need to consider both the risk criticality and the resource needed for addressing the problems. In both cases, planning would be easy because every action would produce the desired planned outcome with the planned resource.

Such a workflow would ignore the real nature of the transport system as a complex adaptive system, a system of systems and a complex socio-technical and political system. As stated in requirement (RT-4), risk treatment needs to embrace the typical features of such a system. The responses of the system to an intervention are unpredictable – even for experts. The best ways to achieve the desired changes are not easy to find and may be counterintuitive. Due to lack of repeatability, interventions which were successful in the past, may fail in the future and vice versa.

Within this complex system, there may be pockets of less complex activities. Therefore, one of the first steps in decision making is to try to understand the phenomenon under study, especially in terms of its level of complexity. This step, reflected in requirement (RT-3) is covered in Chapter 7.5.1.

The general idea is to use the risk picture(s) with the existing background information as the starting point, and to formulate actions and/or experiments, while trying to address all the different, and often conflicting, priorities (e.g. safety, cost, resource, smooth flow of traffic, etc.). One has to accept that most of the time one would be working on estimates rather than known facts. According requirement (RT-5), an adaptive approach should be adopted for interventions. One interesting aspect of the adaptive approach through experiments is that estimates get tested before full-scale implementation. However, this leads to *continuous management of portfolios of experiments*. This does not exclude straightforward actions sometimes, and in some cases experimentation is not even possible.

In the current context, an *action* is a single measure or a plan consisting of series of measures aiming at the desired outcome. An action is designed in a feed-forward manner: there are no self-correcting or

adaptive features within the plan. In a nutshell, experts are trusted to know today, what is the right solution for tomorrow, and what the consequences of the intervention are. The implicit assumption is that the action will succeed in delivering the desired outcome with high probability, and without major side-effects.

The paradigm behind *experiments* (or a *probes*) is different: the unpredictability of the system is acknowledged and the failure probability of any one experiment is considered relatively high. Different experiments are set up in parallel to find the successful ones, which can then be expanded.

A compromise between a simple action and a portfolio of experiments would be an *adaptive policy* (or another type of adaptive solution) which has in-built mechanisms to adapt to at least some future conditions so that the desired outcomes are maintained.

The following sections describe the different aspects of setting up and running the adaptive interaction with the target system - which in this case is the transport system.

5.1 Understanding the phenomenon under study

The requirement (RT-3) states that interventions with the system need to be adapted to the nature of the system. When starting to prepare the definitions of actions and experiments for a specific risk, the first step is to gain more understanding on the phenomenon in question. There are four easy steps for achieving this within the described process. First, at the data integration / knowledge building stage every risk element in the risk picture has been discussed and the different views and conclusions of the discussions have been recorded. Reviewing such integrated information is a good introduction to the topic. Second, the analysis step of this risk element within the risk picture has provided further aspects like the position of the element within the risk picture(s), the kinds of safety events related to this threat and possibly the safety factors that have been involved. The third step is to place the phenomenon into the *Cynefin* framework and the fourth one is to see it from the perspective of the larger systems embedding the local system.

Using the Cynefin framework gives a good reminder of what type of system one is dealing with and therefore what kind of strategies are the most recommendable. In the context of the transport system, in most cases one would land on the *complex* domain. However, it is plausible that some solutions reside in the other domains - for example some technical improvements may belong to the complicated domain. The resulting area in the Cynefin framework produces a very clear guidance on the type of solutions that could be envisaged: there's a big difference between calling for an expert group vs. setting up experiments. If it turns out to be difficult to place the phenomenon in the Cynefin framework the safest choice is to assume the *complex* domain.

A fourth step is to try to understand how the phenomenon makes sense in the context of the larger system(s) which host the local system, i.e. applying *synthesis* instead of analysis. For example, local pressures often have their "roots" within the larger system, which may require higher efficiencies, and which get passed on to the operational level through the local management. At the system level it may also sometimes be useful to draft some kind of *system diagram*. Already identifying the key interactions between different actors in the system can be very useful. Further study may add in power structures and influencing channels. Even a rough model may help in anticipating some potential consequences of certain interventions, or help identify promising leverage points.

Due to the cyclic nature of the risk management process, there may also be results available from previous experiments. Such aspects are covered below in the context of the NRMF as a process.

5.2 Designing experiments

As discussed above, it can be expected that most emerging safety issues will belong to the complex domain. Even technical issues which as such would belong to the complicated domain usually interact

strongly with the human system thereby bringing the overall issue in the complex domain. As stated in requirement (RT-4), risk treatment intervention need to embrace the CAS-nature of the system.

Using experiments brings with it a new mindset: an iterative culture where the aim is a gradual convergence towards successful solutions through experiments, rather than trying to be able to define the winning solution at one shot.

The first step in setting up a portfolio of experiments is to try to make sure that the portfolio is rich enough and features also unusual and disruptive ideas. For this purpose, the group of people proposing the experiments should be diverse enough. Subject matter experts are not necessarily the people who propose novel solutions. The group could include ordinary professionals who are dealing with the system in question in their work (e.g. crew members) but also complete strangers to the activity or people who may have faced similar problems in a completely different industry. It is an advantage to feature different educational backgrounds, i.e. not only technical and operational people.

The experiments could aim to solve the main issue directly or they could try to solve a related problem. Sometimes the latter strategy may eventually lead to a good solution to the main problem - in the same way that coming up with a superior battery technology enabled the current generation of mobile phones and could well play a key role for electric vehicles and even electric aircraft. Because the expectation is that most of the experiments fail, they must be *safe to fail*. In other words, the failure of an experiment should not come with excessive cost or other excessive negative consequence. In fact, if most experiments do not fail, it could mean that the potential solutions have not been explored widely enough. This also means that proposals for experiments should not have to go through a very strict screening process, as long as they are safe-to-fail and generally coherent.

Each experiment should come with a clear definition of what would constitute success/failure and how the experiment could be enlarged if it was successful and faded away it was a failure. The experiment should be restricted in place and time, for example use a small town as a testing ground for new rules affecting young drivers and run the test for 3-6 months.

Once the experiments are running their results would be followed and failing experiments would be stopped while the successful ones could be enlarged. Some experiments could also be combined or redefined. As a result of this iterative process, one should obtain a solution which brings the desired outcomes in an adapted manner and minimizes unwanted side effects.

As mentioned in the previous section, there are some limitations to when experiments can be used. One challenging area is the black swan corner due to the low probabilities and the high number of scenarios. For example, it would be easier to define experiments against young people driving drunk after parties during the weekend than against a threat of a commercial airliner accident at takeoff due to wrong weight data inserted in the flight management computer. The latter would consist of a high number of potential scenarios involving very rare errors or violations – so it could be very difficult to experiment with strategies against them. Interacting with the transport system in relation to risks placed in the black Swan corner are covered in Chapter 7.5.5.

5.3 Defining actions

There are several reasons why an action may sometimes be preferred compared to experiments. First of all, the notion of safe-to-fail experiments along with the acknowledgment of the unpredictability of a complex system is a new challenging paradigm for today's decision-makers. The status quo for a long time has been to address problems with actions, defined by a group of respected senior managers or experts. An action is also often simpler to set up and to follow up (if any follow-up is done at all) compared to defining and running a portfolio of experiments in parallel and managing them dynamically. There are also situations where experiments may be difficult to implement even when desired. In a highly standardized environment (e.g. standard operating procedures) it may be difficult to change the procedures locally even for a limited period of time because it fights against the idea of

standardization and could increase operational risks. It may also be difficult to measure the success of an experiment, e.g. if the system is already very safe.

At the level of a national safety agency, actions aimed for risk reduction could typically consist of new regulation, limiting or forbidding certain types of operation, requiring additional training or increasing oversight on certain operational areas. At the level of an operator, typical actions could be change of operational procedures, amending the contents of training programs or increasing restrictions on certain parts of the operation. At both levels, actions could also consist of engaging in improvement projects with third parties like international organizations or equipment manufacturers. It is argued that more imagination could be used in defining the types of actions that state agencies launch. The case study in in Part III contains examples of possible action types, including some less conventional ones.

When an action is defined, it is important to make the underlying assumptions explicit and describe the expected outcomes of the action. This helps following up the effectiveness of the action and gaining lessons learned for the future.

5.4 Designing adaptive policies

A compromise between experiments and simple actions would be to use adaptive solutions, such as adaptive policies. By definition, one would expect such policies to be more successful in dynamic conditions than non-adaptive policies. The basic idea is to make the policies conditional in such a way that foreseeable – and ideally even unforeseen - evolutions can be accommodated.

Sometimes the policy would update itself automatically (e.g. a value would change in relation to an outside indicator), and sometimes the policy would simply flag up a need for revision based on certain conditions having been fulfilled. In any case, such policies require follow-up.

It is easy to say that adaptive policies can be recommended compared to non-adaptive policies. However, it is easy to understand that such policies are more difficult to design and the tradition and know-how related to such work is limited. The dynamic nature of these policies could be seen to increase uncertainty within the community as the fact that the content will change to a yet unknown form, is very visible. If the design is not carefully done, vulnerability to unwanted consequences could increase, as well as deliberate abuse or gaming with the in-built mechanisms of the policy.

5.5 Treating risks in the black swan corner

In terms of trying to manage the risks of the top right corner, one is dealing with scenarios with very low probabilities, possibly involving incredible coincidences and extremely improbable combinations of factors and occurrences, so the number of different scenarios could easily rise to an unmanageable figure. One cannot think of all possible scenarios, so one cannot “catch” the black swans. Even for identified threats, there are potentially thousands of possible related scenarios with probabilities that are unknown. Due to this nature of the scenarios, the situation is also bound to be more volatile and more prone to coincidences than in the other areas of the risk picture. Here, small details may have big impacts in terms of the outcome. All this means, that trying to address these threats systematically, scenario by scenario, is a desperate strategy. As argued in Chapter 2.3 and reflected in requirement (RT-2), the strategy for dealing with this type of threats must rely on solutions which increase the resilience of the system rather than trying to address individual scenarios. More concretely, the operational crews and the organizations around them must be ready to encounter surprises, potentially taking them outside of the trained operational envelope. Enhanced resilience could help lowering the risks both for known and unknown threats.

As an example, a crew controlling a big cruise ship was observed to control the ship in a semi-manual steering mode despite a fully automatic mode being available (see Chapter 9.2). This was done in order to maintain the manual control skills of the crew even though these skills are not normally needed. This is a good example of building resilience. The ship could get into trouble through thousands of different scenarios but it could be envisaged that in many of them one of the common factors would be the need

to be able to control the ship with less available automation than during the normal operation. Therefore, rehearsing these manual modes provided readiness for a multitude of scenarios at least in the sense that the additional challenge of manual steering becomes much less of a challenge. In Woods's (2015) terms, the graceful extensibility is enhanced.

The strategy of increasing resilience can be implemented both at the level of operational crews and at the level of the whole organization. At the operational level, a key focus area will be in assuring that all *operational competencies* are at a high level both for normal operation and abnormal situations. The approach needs to cover both knowledge and practical experience practicing different situations. Often this is not possible as a part of the normal operation so dedicated simulation facilities can be used. Even if every simulation session needs to use specific scenarios for conveying the training objectives, the overall focus should not be on the given scenarios as such but on *creating generic competencies, skills and knowledge* which will be useful in situations far beyond the few simulated cases. *In the context of building operational resilience, the safety factors can be of great help*. The crew competencies within the safety factors provide a good checklist for making sure that all the necessary competencies are developed enough. The rest of the safety factors, and especially the *fundamental safety factors*, point to the critical factors that are counted on for safe operation, and thereby propose focus areas for building more resilience around these critical points.

At the level of the whole organization, building resilience can focus both on normal and crisis operation. In the normal operation, it is important that operationally critical functions have a sufficient level of safety margin. Organizational resilience at the crisis mode covers many things that can be put in place in advance. For example (adapted from Woods 2014):

- The normal hierarchic *decision-making* process is usually too slow for a crisis situation. Therefore, it needs to be clear that operational staff have the *authority* to take all the necessary decisions during the crisis situation without hesitation on whether they should get acceptance from higher management levels.
- It needs to be clear *what can be sacrificed*: for example, it may be that the normal financial expense limits can be exceeded in a crisis situation to ensure the best possible operational outcome.
- The standard procedures may not be applicable when the operation has moved outside the normal domain. Therefore, the staff needs to know that *taking initiative* can be a crucial skill during the crisis - even if it was strongly discouraged during normal operation.
- The use of different *communication channels* can also be practiced and various *difficulties anticipated* and covered in training.
- The normal operations usually optimize in terms of resources, so any exceptional situation would cause a situation where extra resource is needed rapidly. Mechanisms can be put in place to enable *obtaining extra resource rapidly* in a crisis situation.

Obviously, these are just examples to illustrate the multitude of ways in which resilience can be built at the organizational level already during the normal operation. Addressing the top right corner involves developing operational and organizational resilience - and inversely, when one wants to build resilience, it can be very useful to review the threats in the top right corner because they can give vivid images of the kinds of situations that one might have to deal with one day in the future.

Other possible strategies for dealing with threats at the black swan corner include:

- Capturing ways in which operational crews are creating resilience and reinforcing and spreading such practices.
- Creating and using feedback to detect deteriorating resilience
- Developing ways to rely on simple autonomous artifacts (either as the normal or the backup solution), for example using whiteboards vs. computers dependent on software and power.
- Rehearsing surprising situations and demanding escalating scenarios.

The generic resilience-enhancing strategies may often conflict with other priorities. Having sufficient resource for surprise situations would typically mean having a more expensive resource structure during normal operation. The notion of promoting self-initiative during crisis situations while asking for strict adherence to standard operating procedures during normal operation can be difficult for many people. Imagining extreme scenarios and rehearsing them through simulations could seem too expensive for some people, especially if there are no major accidents in the recent history.

Finally, it is worth noting that not all resilience-enhancing measures have to do with people. It may be possible to introduce technical solutions which create additional resilience against many different threats. As an example, the structural split of a ship into watertight compartments enhances its resilience against many kinds of scenarios where the integrity of the hull is lost and water starts flowing in.

5.6 The second level: organizations

So far, the focus has been mainly on threats and other elements that are presented in the risk picture, and addressing them as such, independent of the organizations in which they emerge. This is the first level at which interaction with the transport system can take place. The second level is at the level of organizations.

If the perspective of a national safety agencies adopted, the key organizations which carry operational risks are the operators. Other relevant organizations would include national federations for transport-related hobbies like sailing, car sports, gliding and general aviation.

Interacting with these organizations is at least as important as focusing on individual scenarios. Often dealing with an individual scenario would eventually involve interactions between organizations. Additionally, as was seen in the previous section, at the organizational level it is possible to develop additional resilience which can be applicable to multiple threats and scenarios. The focus in this section is the latter: how to estimate the vulnerability/resilience of various organizations and how to try to improve the transport system at that level.

Ideally, it should be possible to monitor the organizational safety levels of different organizations and launch adaptive responses. A safety agency would be most interested in the operators and would try to understand and enhance their safety levels. An individual operator would focus on different parts of its operation as well as on its external service providers. This could also include different segments of the operational crews etc.

In practice, the task is challenging. The available data in itself may not allow drawing reliable conclusions on the organizational safety levels even if some events may give tangible hints in one direction or the other. It is also politically touchy for a safety agency to start classifying operators in different “safety categories”: this could be interpreted as a safety ranking list and therefore the operators could insist on having hard evidence to justify the classification. The evidence would always be subject to interpretation.

Ironically, the most reliable feeling about the organizational safety of an operator is probably the luxury of principal inspectors in domains where a single inspector has a long-term relationship with an operator and gets detailed insights to the operator through audits, visits and many other interactions. This would be the case in civil aviation. The irony comes from the fact that the understanding of the organization in question is not based mainly on hard data (e.g. safety events) but rather on the subjective experience of a domain expert having a long-term exposure to the organization. The audit findings as such could be used as hard data but they only provide a partial view. The inspector knows more than can be read in the audit findings. That is also why such people are absolutely invaluable in the data integration stage of the process to build the strength of knowledge.

There are attempts to track some key aspects of organizations typically in the form of a table, assign points for different categories and use the sum of the points as an indication of the safety level or

resilience level of the organization. It is argued that compared to the real complexity of the organization's and the nature of safety is an emergent property such tables are too simplistic for providing a reliable understanding of the true resilience level. In any case, safety agencies in several countries run a continual process where operators' safety levels are estimated using such simple approaches. The results are typically used to guide the oversight activity. An example of such a point-scoring method used by the Swedish Transport Agency can be found at Transportstyrelsen (2014) and a similar method used by the Finnish Transport Safety Agency is discussed in the case study of Part III.

The conclusion is that safety at the organizational level is too important to be ignored but realistically speaking too difficult to measure in a way which could be fully quantified and justifiable. The pragmatic compromise would therefore be to use all the available information and integrate it together in the same way as for threats. The objective is to form a qualitative picture of the organizations. The inputs from people interacting a lot with these organizations are of primary importance even if their information could often be in the form of stories rather than quantifiable data. Interactions with the organizations also give the opportunity to discuss the perception of the agency directly with the organization in question. Such discussions with the top management and with corporate safety staff can be very revealing and useful for both parties.

The theoretical base for safety at the organizational level was developed in Part I. Based on that, a list of focus areas for safety at the organizational level could be created. These could be called *organizational (level) safety factors*. Even more than the list of operational safety factors, this list should be considered a very approximative product aimed for giving rough guidance on focus areas for learning about organizational safety and resilience.

The HRO factors from Sutcliffe (2011) provide a good starting point to organizational safety/resilience. The fourth point on resilience can be replaced by the points on resilience [3] and [4] as proposed by Woods (2015):

- *Preoccupation with failure*: is chronic awareness of risks and potential hazardous events maintained?
- *Reluctance to simplify interpretations*: are different opinions and alternative perspectives encouraged?
- *Sensitivity to operations*: is there enough interaction and information sharing at the operational level to maintain an integrated big picture and to share talents and skills?
- *Graceful extensibility*: can the system *stretch* to handle a challenging surprise?
- *Sustained adaptability*: does the organization demonstrate sustained adaptability over time?
- *Deference to expertise*: during high tempo situations, are decisions delegated to the people with the most relevant expertise, irrespective of authority or rank?

In order to cover the basic operational capabilities, the following questions are proposed:

- Are the core operational competencies mastered in the organization? Has maintaining those skills been ensured also for the future?
- Does the organizational structure adequately support the operation?
- Do the processes function properly?
- Are the equipment, tools and facilities adequate?
- Are there enough resources in the organization at each level? Can resources be increased fast enough to adapt to higher demand?
- Is there a good atmosphere and spirit among the people? Are people engaged and motivated?

Concerning safety management in a complex environment:

- Are safety matters prioritized high enough compared to other priorities - at all organizational levels? Is there enough power behind safety?
- Does the safety management system function properly?
- Is the safety work in the organization based on modern understanding of safety and risk (e.g. safety as an emergent property)?

- Is there an understanding of phenomena affecting complex systems, e.g. drift, normalization of deviance?

The capability of an organization to get an idea of the organizational safety of another organization depends a lot on the access to the relevant information. Even when the access is not a problem, another natural limitation is the time and resource needed to interact and get to the relevant information. As much as the organizational safety is an important matter, the actual implementation for tracking it will always be a compromise between these factors.

Despite the challenges related to understanding the safety level of an external organization, it is proposed that such activity is an integral and continual part of the risk management process. The organizational safety level would be discussed in a similar way to discussing threats, and the integrated understanding would be recorded for future reference. The tracking could be purely qualitative or it could include quantitative elements. In both cases, it is recommended to record not only the result but also the related discussions and conclusions.

5.7 Combining analyses on threats and organizations

Once both threats and organizational safety levels are being addressed within the risk management process, it is possible to combine the results from the two and obtain an even more precise targeting of resources on the most critical points in the system. This resonates with one of the fundamental objectives of risk management: allocate resources in such a way that they have the best possible impact on the risks. For example, the may be a few high criticality threats; but maybe these are not applicable to all the operators due to their different operations. So already some resource can be saved as some of the operators can be left out from the scope. Among the concerned operators, maybe some are considered very resilient so they can again be trusted to manage the threats in question. Finally, the real scope for active treatment of these threats could be the operators who are exposed to this threat *and* are not considered very resilient at the moment.

How to carry out this cross analysis in practice? The threats are in the risk picture but the information on safety levels of the different organizations is probably in some sort of qualitative description of each organization - potentially with some simple classification. One way is to build a two-dimensional table which allows to focus on each intersection between different threats and different organizations case-by-case. For example, it could be helpful to sort the organizations in the table roughly by the level of perceived resilience. The applicability level of different threats to different organizations could be checked first, crossing out the squares which become non-applicable in this way. The remaining exposed organizations could then be considered for their resilience and the most critical combinations could be identified.

Again, a software tool could be helpful for this step. It is also possible to try to gain more depth in this analysis by looking at the safety events: which organizations experience the events? How well do they manage? What are the related threats? The operational safety factors can also be used to study how much an organization might be exposed to a specific threat.

Such a cross-analysis could in itself be the basis for defining specific actions and/or experiments. It could also give decision support in relation to already planned actions and experiments.

5.8 The third level: the transport system

As mentioned above, the third level at which interventions can be studied is the level of the whole transport system. One can for example ask, why do operators tend to drift into an operational mode where resilience is sacrificed to the benefit of increasing operational efficiency for the normal everyday operation. Such questions can be answered properly only by taking the system-level perspective.

System-level problems can only be addressed by system-level solutions. In many cases solutions at the level of threats or organizations (levels one and two) would not work, or would only have a short term

impact and then gradually fade away due to system-level influences. For example, forcing an operator to add some extra resources to manage a safety critical task may be a big help in the short term. However, as the whole organization is subject to the same operational pressures, gradually the new resources could become consumed more and more by various operational duties. The extra resources may also indirectly lead to increasing the operational volume. Consequently, there will be finally less and less time for the new resources to really focus on the safety critical task they were originally assigned for.

This does not mean that interventions at levels one and two are not useful. Study of the problem through the Cynefin and possibly a draft system model may give strong hints when intervention at system level is a must.

The guidance on dealing with the complex adaptive transport system was summarized in Chapter 3.5.4. Perhaps the most valuable guidance comes in the form of the hierarchy of interventions. Impact can be very significant if one can have an influence at the level of the system's *purpose* (or function). Here one needs to keep in mind that the real functional purpose of the system is not necessarily the purpose that is communicated (or that people believe in). Leverage can also be relatively high if one is able to change the *interconnections and relationships* within the system. Simply changing *elements* within the system, like individuals, is a much less promising course of action.

Another useful reference is the complex leadership concept (Marion & Uhl-Bien 2001), which could be applied at the system level. The overarching theme is facilitating and catalyzing beneficial emergence of networks and largely self-organized innovation.

What could such strategies mean in practice? Operators may have leverage on many matters especially if they act together. They can for instance create new standards by working together. International organizations and associations can often be the platforms where agencies and operators and other stakeholders can come together and try to create more profound changes in the transport system. In some cases, even individual professionals may be able to start a movement which creates a significant change in the system. A good example is the creation of the Evidence Based Training EBT concept in civil aviation. An experienced pilot instructor started a movement to change the paradigm behind pilot training from focusing on scenarios to focusing on key competencies. The idea was also to study through available data what should really be the content of the training instead of repeating some very old training syllabi. By engaging a group of instructors around the world and through organizations such as ICAO, IATA and IFALPA, EBT became a referenced ICAO training concept. EASA is currently developing the related regulatory framework. Such an example shows the positive side of the non-linearity of a complex system.

Even when the interventions as such would be aimed at levels one or two, it is always important to remain aware of the whole system and its complex nature. This includes acknowledging that many behaviors and phenomena are actually defined by the bigger system rather than the local context. For example, a CEO prioritizing business objectives compared to safety objectives may be seen as the key problem. However, the larger system around the CEO defines the dynamics which have a huge influence on his/her behavior. The company shareholders want their earnings. The competition creates a constant pressure on prices and therefore on revenues. The incentive structure for the CEO typically has a significant variable part which may be mainly driven by the short-term financial performance of the company. The safety level is not part of the incentives and even if that was desired it would be difficult to implement in some kind of measurable way. These and other factors create a situation where the CEO is strongly pushed towards a certain model of operation - no matter who he/she is. It is difficult to fight such a systemic phenomenon at the level of the CEO or below. Here the system in question is not even the transport system or a part of it. It is the world economy as we know it today with its financial and political dimensions. Adopting a systems focus will be a prerequisite for finding effective leverages.

The complex nature of the system also means that planned interventions cannot be considered one by one, in isolation from each other – as reflected in requirement (RT-6). *Actions/experiments will often have direct or indirect interactions between themselves*. There may be both synergies and negative

effects. For example, certain operational restrictions could each have a minor negative business impact on operators if only one of the restrictions is imposed. However, the combined effect of two restrictions could be devastating enough to put some operators out of the business, due to the dynamics between the two restrictions. Even if foreseeing the interactions can be very difficult, there should nevertheless be an attempt to study the *combination* of all interventions every time new ones are added or old ones are removed. Small-scale experiments can again be very useful in highlighting interactions.

Balancing different priorities in a consistent manner reflecting the recognized values is another challenge. Proposed "desired changes" and interventions should be systematically challenged in relation to all the different priorities to find the acceptable balance. How low can speed restrictions on roads be, without driving becoming too frustrating and time-taking? The case study in Part III presents one way to challenge interventions from different perspectives, using the *ritual dissent* method.

6 The risk management framework as a continual cyclic process

The above sections have described different parts of the overall risk management framework. This section focuses on how to operate the framework as a continual cyclic process. At any point of time, there will be a number of issues in progress within the different sub-processes. There are ongoing experiments, new threats to be identified, the risk picture to be updated, knowledge to build, issues pending decisions and so on.

The repetitive stages in the process would include at least the following:

1. Identifying new threats, safety issues and scenarios
2. Integrating all available information and knowledge about a threat to enhance the strength of knowledge
3. Placing the new threats and safety issues in the risk picture
4. Studying the risk picture as a whole and identifying priorities
5. Monitoring relevant organizations and building knowledge about them. This could include organizational safety factors and safety/resilience profiles.
6. Combining the perspectives on threats and organizations to carry out cross-analyses
7. Drafting experiments, actions and adaptive interactions (e.g. adaptive policies)
8. Overseeing and following up the implementation of the above
9. Monitoring the impact and success of the various interventions
10. Taking decisions on new interventions and on adapting the existing ones (including discontinuing some)
11. Carrying out data processing and preliminary analyses to support the various steps in the process (e.g. identification of new threats, knowledge building, intervention selection).
12. Integrating results from each mode of transport into a combined risk picture and carrying out all the steps of analysis, intervention drafting, decision making, monitoring, follow-up, etc. at this multi-modal level.

A natural balance could emerge between the modes of transport individually and the combined, multi-modal process. Each mode of transport could run its process based on its own specific risk picture fairly independently, but the content from each mode would be regularly reviewed at the level of the multi-modal risk picture. The multi-modal perspective provides:

- The obvious benefit of *optimizing interventions at the level of the whole transport system*
- An opportunity for *constructive challenges* between different modes of transport (e.g. if there are significant differences in cost-benefit values of interventions between the modes).
- An opportunity for *sharing lessons learned*, e.g. on the success of experiments.

In an extreme case, all decision-making could be done solely at the multi-modal level, in which case all interventions at the level of the modes would be subject to acceptance at the multi-modal level. While in some ways such a process would be the purest form of trying to optimize the interactions with the

transport system, it would probably not be the ideal case. Decisions would move further away from the people with the knowledge and experience and/or the risk workshops and other meetings would grow very big in terms of participants and duration, with excessively long agendas. This would make the whole process more bureaucratic and less attractive for the participants. On the other hand, if the multi-modal level was ignored and all decisions were taken only at the level of transport modes, the opportunities provided by the multi-modal perspective listed above would be totally missed.

The listed process stages would involve three types of organizational bodies:

- Core of the process would be assured by the so-called *risk workshops*
- An *analysis function* would be needed to prepare the risk workshops
- A *decision-making body* would be necessary to validate all significant decisions, with the commitment for the associated resources. There could be a large overlap between the participants of the risk workshops and the decision-making body.

For a multi-modal agency, the risk workshops and decision-making bodies could exist for each mode of transport and at the multi-modal level. A single analysis function supporting all the bodies makes sense due to the specific skill and tool requirements, and the fact that such team would probably already naturally exist in the organization. Most probably, the risk workshops and decision-making bodies would not be permanent organizational units, whereas the analysis function would.

The stage 5 about monitoring relevant organizations could be carried out by the oversight function, the analysis function or the risk workshops. The cross-analyses (e.g. threats vs. organizations) require involvement of the analysis function. The results should be used by the risk workshops.

The following sections cover the functioning of the risk workshops, decision making bodies and the analysis function (in relation to the risk management process).

6.1 Risk workshops

The members of the risk workshops consist of some permanent members and some invited members whose presence depends on the topics covered. The permanent membership has the function to ensure bringing in the necessary knowledge and information on the threats, organizations and the different parts of the transport system. The expertise is reinforced ad hoc by inviting subject matter experts guided by the topics covered each time. Some of these experts may be external, typically from another agency or an operator. Finally, it is also recommendable to include invited members who are not necessarily experts but ordinary users of the transport system. For example, when the safety of young drivers is discussed, some young drivers could be invited to be present. Such users can expose the reality from their niche within the transport system to the risk workshop, give a reality check to the assumptions made and to criticize the proposed interventions and to contribute to their designs. Whenever interventions are discussed it is also important to have people present who can defend other priorities such as environmental sustainability so that the safety-driven interventions do not sacrifice other priorities too heavily.

A risk workshop (for one mode of transport) would typically ensure the stages 1-9 in the above list. Due to the number of tasks, risk workshops need to be organized regularly and rather frequently.

There are obviously many different ways to organize the agendas of these workshops. Some sessions could be dedicated to a single topic or all meetings could cover different topics and different stages of the process. An example agenda could look like this:

1. *Introducing* new threats and safety issues, if any. This includes work on potential surprises.
2. Focusing on one or more selected threats/safety issues. *Building knowledge*. Placing/updating these elements in the risk picture.
3. *Studying the risk picture* as a whole, both in terms of risks and potential interventions.
4. Studying and discussing the *actions and experiments in progress*. Recording the *observations*.

5. Discussing whether there are any *observed or expected changes, trends or phenomena* in the transport system which could have an impact on transport safety and should therefore be taken into account.
6. Making *conclusions* based on points 3-5 and recording them.
7. Drafting intervention *proposals* for new elements (from points 1-2), and for adjusting the ongoing interventions.
8. *Selecting* the threats/safety issues to be focused on in the next meeting.
9. Any other related issues.

It is important that the topics are selected before the workshop so that participants can collect the necessary information on their side and reflect upon it in advance. The sub-agenda for processing a single threat/safety issue could be the following:

1. If the analysis function has done preparatory work on the topic, they might be in the best position to give the first overall introduction to the topic.
2. A round-the-table could start with people who have a good overall view to the topic. For example, this could be a principal inspector if the topic relates to an operator. In this phase of sharing information, the objective is to bring the different views and all available information to the use of the whole group. The danger is starting to make conclusions too early with partial information. Therefore, discussion could be kept to the minimum. Clarifying questions can be encouraged. A Delphi-inspired first stage of purely individual inputs could also be used.
3. The following data/information needs to be reviewed. Parts may be covered before/during/after the round-the-table, as seen fit.
    a. Review of the data/information gathered by the analysis function on the topic. This can cover information on the relevant threats, scenarios, safety events, safety factors and so on.
    b. Results related to relevant oversight information, e.g. audit results.
    c. Relevant information/knowledge that the participants possess.
    d. Relevant research findings.
    e. An acknowledgment of any blind spots. It is good to create a shared awareness on any aspects that would not be visible with the current information sources. This supports recognizing uncertainties.
4. Open discussion. A good principle is to make sure that all the participants contribute to the discussion and that their contributions are recorded. This ensures taking into account all the diverse perspectives to the matter. Everybody has the responsibility to speak aloud, especially when there are disagreeing perspectives. Diversity needs to be nurtured.
5. Identification of the different scenarios associated with the threat.
6. Review of uncertainties and assumptions. A separate thought may be given to how the situation may develop in the future.
7. Making a synthesis. The synthesis should capture all the relevant discussions and the conclusions on the matter, acknowledging the uncertainties and without trying to simplify the situation too much. There is no need to reach consensus. Recording the differing views may be even more useful than a single shared view to the matter, because it supports the design of parallel experiments.

The given agendas are just examples but they illustrate how the different processes can be managed in parallel through regular risk workshops. At least the key topics discussed and all the outcomes need to be recorded for future reference. This can be done in innovative ways, e.g. filming person(s) how summarize the topic(s) verbally, rather than typing everything.

It is obvious that through all the tasks that the risk workshops carry out, they possess a better understanding of the various risk elements than anyone else. Therefore, it would be logical that proposals for interventions also were prepared by the workshops. However, other arrangements are possible too. For example, it may be considered that the detailed design of the interventions requires too much time from the workshops and/or specific knowledge that can be found outside the risk workshops. In that case, the job of designing the interventions could be delegated to one of several task forces (possibly

from existing organizational units) specialized on different topics, or shared between the risk workshop and a task force.

An important step in the design of the interventions is to challenge them from the perspective of the various (often conflicting) priorities. For example, what is the influence of the (safety) intervention on the availability/reliability of transport services, on the environment and on the businesses providing the transport services? One way to arrange such a challenge is to assign knowledgeable small teams to challenge the proposed intervention from the perspective of all the key priorities. This can be done as a relatively quick live exercise (during the intervention design phase) by the designers collecting feedback from all the assigned "challenge teams" and re-designing the intervention enough times to reach a workable compromise. This type of challenging is less relevant for experiment because they could be adapted quickly if conflicts with any of the priorities emerge.

Once actions, adaptive policies and experiments have been launched, managing the related portfolios becomes a continual activity. As the concrete implementation most probably falls on permanent organizational units, the portfolio management may become a shared duty between the implementing organizational units and the risk workshops, even if the latter need to maintain an overall understanding of the interventions.

6.2 The support role of the analysis function

The analysis function will have a key support role in the process. The risk workshops are challenging in terms of time because there are a lot of matters to discuss. Therefore, all possible preparatory tasks should be done between the workshops. Such tasks can naturally be assured by an existing analysis function in the organization.

The preparatory tasks would include at least:

- Proposing the agendas and the specific topics of focus.
- Assuring the analysis functions related to incoming data, including categorization and preliminary analyses, even risk assessment and application of safety factors.
- Carrying out risk identification and proposing new potential threats and safety issues to the risk workshop.
- Proposing a preliminary set of scenarios related to a new threat/safety issue.
- Proposing a preliminary position, size and uncertainty value for a new risk element to be placed in the risk picture.
- Maintaining the risk pictures.
- Maintaining the links between safety events and the other elements in the risk picture.
- Helping the risk workshop to make sense of the risk picture by proposing appropriate ways to visualize it.
- Integrating the cost aspects into the risk picture.
- Gathering information related to a safety issue on the agenda.
- Identifying the blind spots related to a threat or safety issue.
- Supporting the monitoring of ongoing interventions through available data.
- Integrating elements from the risk pictures of each mode of transport and constructing the aggregated multi-modal risk picture.
- Assuring the same preparatory tasks both for each mode of transport separately and for the common multi-modal process.

Additionally, if the analysis function is already the owner of safety data, it may be practical that it also becomes the owner of the information related to the risk workshops, such as the records of discussions and conclusions.

6.3 Decision making

One of the key concerns related to decision making is that the decision-makers should have a detailed understanding of the topic, including all the background discussions, analyses, assessments and shared expert views. In this respect, the ideal case would be that the decisions are taken within the risk workshops. However, this may introduce other types of problems. The experts in the risk workshop may not have the organizational positions that authorize them to take the necessary decisions. Additionally, if the organizational units which will have to implement the decisions are not involved in the decision-making, they would probably not feel convinced and committed to implement the interventions. In line with previous discussions about the role of humans and the importance of representing information within its context, the decision-making should not be completely separate from the risk workshops. Yet another consideration is that people who have high organizational positions may have too much influence in the risk workshops and increase the tendency for groupthink.

Consequently, the ideal set up could be an overlap between the risk workshops and the decision-making bodies. The overlap would ensure a sufficient information transfer between these two bodies. The highest-level decision-makers in the organization should probably not be part of the risk workshops, or participate only during some stages. The latter option would have the benefit that these decision-makers get exposed to the details behind the proposals.

Ideally, the intervention proposals would come from the risk workshops and the implementation would be as close to the original proposal as possible.

If each mode of transport runs its own risk workshops, the most optimized arrangement could be one where these workshops are taking place roughly at the same time and are being followed by a multi-modal risk workshop where the common risk picture is reviewed together. This global review could be followed by decision-making at this multimodal level. The local transport mode level decision-making meetings could then follow and take benefit of the higher-level decisions. At the multi-modal level enough participants from the local workshops are required so that all the necessary information is present in this group. As all the preparatory and assessment work has already been done at the local workshops, the multi-modal workshop can concentrate on the global review of the risk elements and intervention portfolios.

6.4 The role of humans in the process

The aim throughout the development of the risk management framework has been to see people within the process as key contributors to good results, and to build on their strengths - rather than trying to design a human-free process or to try to limit the role of humans to the minimum. The ways in which the positive capabilities of people are embraced during the process include:

- Categorizing incoming safety events properly. This is often not simply a question of keywords but rather understanding the larger context.
- Carrying out event risk assessment and recording the reasoning behind for potential future debates among other experts.
- The first two tasks require interpretation of narratives and understanding the context which is not necessarily part of the narrative.
- Knowledge building at the risk workshops.
- Producing diversity and different perspectives thanks to participants both from inside and outside the agency.
- Making a living link for information and knowledge transfer during the process, e.g. between the risk workshops and the decision-making bodies.
- Making judgments and taking decisions, involving multiple priorities and value judgments.
- Enabling constant learning and building wisdom. This aspect is discussed in the next section.

The requirement (RT-9) promotes taking advantage of these human capabilities but also requires countermeasures against the foreseeable human biases and limitations. The potential traps and the related countermeasures include at least the following:

- Groupthink and defensive behavior: diversity of the members (including external members), some members changing as a function of the topic, not including the highest-level decision-makers in the risk workshops.
- The tendency of a group to act in line with the availability heuristic by leaving the knowledge building stage too early and jumping to conclusions can be counteracted by specifying that no solutions/proposals are allowed during the knowledge building stage. The facilitator can also ensure that a certain minimum time is always spent on the knowledge building and that there is no rush to move on to the next stage if there is still useful knowledge to process.
- Using experiments and specifying in advance what success/failure would look like is a good way to counteract many biases: the truth coming out in the form of the result of the experiment challenges any preconceived ideas (e.g. due to groupthink) and fights against confirmation bias and the tendency to minimize uncertainty.
- Part of the challenges related to understanding complex systems can be fixed with training, which should cover at least the basic features of complex adaptive systems, the Cynefin model and system diagrams.
- Better and better mental models are supported by a conscious effort for long-term learning (see next section), systematic knowledge building and the culture of experimentation.
- It will be easier to adopt a culture where uncertainties can be openly admitted if all the key stakeholders are first educated in this matter. The stakeholders would need to include at least the other agencies, operators, the ministry, and even the representatives of the media.
- Political pressure to use a disproportional amount of resource on a specific topic (e.g. after an accident) may be managed by placing the issue in the risk picture and illustrating its relative importance compared to other risk elements.
- Any unwanted tendency to be risk-averse (against single high-impact accidents compared to several low-impact accidents) can be neutralized by using the severity points.

6.5 Continuous learning and coevolution

As reflected in requirement (RT-8), one of the key objectives of the whole process is the collective learning which takes place throughout the different steps. Ideally, this is not just a question or information or knowledge but learning at the level of wisdom: understanding underlying patterns, updating mental models of the system and its dynamics, and being aware of what the desirable longer-term developments at the system-level are.

More specifically, the different steps in the process provide learning opportunities at least in the following ways:

- Capturing feedback from the system in terms of data and reactions to ongoing experiments, policies and actions
- Collective learning by sharing information and knowledge during the knowledge-building
- Seeing the holistic picture of risks
- Seeing the organizational-level results (resilience, safety factors)
- Drafting and observing system-level phenomena
- Seeing the holistic picture not only based on risk criticality but also including the cost dimension for achieving the desired changes
- Seeing what kind of implementations flow easily and which will prove more difficult in practice

In such a rich context it is difficult to imagine that learning could happen only through documents or software applications. The continuity created by a relatively stable group of people and making sure that new people are gradually introduced in the process when existing ones leave, are assumably important success factors. The most useful resource for knowledge and wisdom will probably reside within the people rather than in the records. It is also the reason why it is recommended that when information

flows between steps in the process is always accompanied by knowledgeable people. The people involved can apply the lessons learned immediately in the following cycles of risk management.

Taking a larger system perspective, a continuous coevolution should emerge, involving the agencies, the operators and potentially even the customers of the transport system. The organizations adapt to the changing system, and the system evolution is influenced by the activities of the organizations. The learning and the coevolution along with the desired and achieved system improvements are really the ultimate goals and deliverables of the whole risk management framework and process. In a constantly evolving system, the goal cannot be a predefined state but rather a constant learning process where the risk management process stays sufficiently ahead of the game in order to keep the risks at sustainable levels.

6.6 The risk management process summarized

The proposed risk management process is presented in Figure 25, using the basic structure of the ISO 31000 framework.

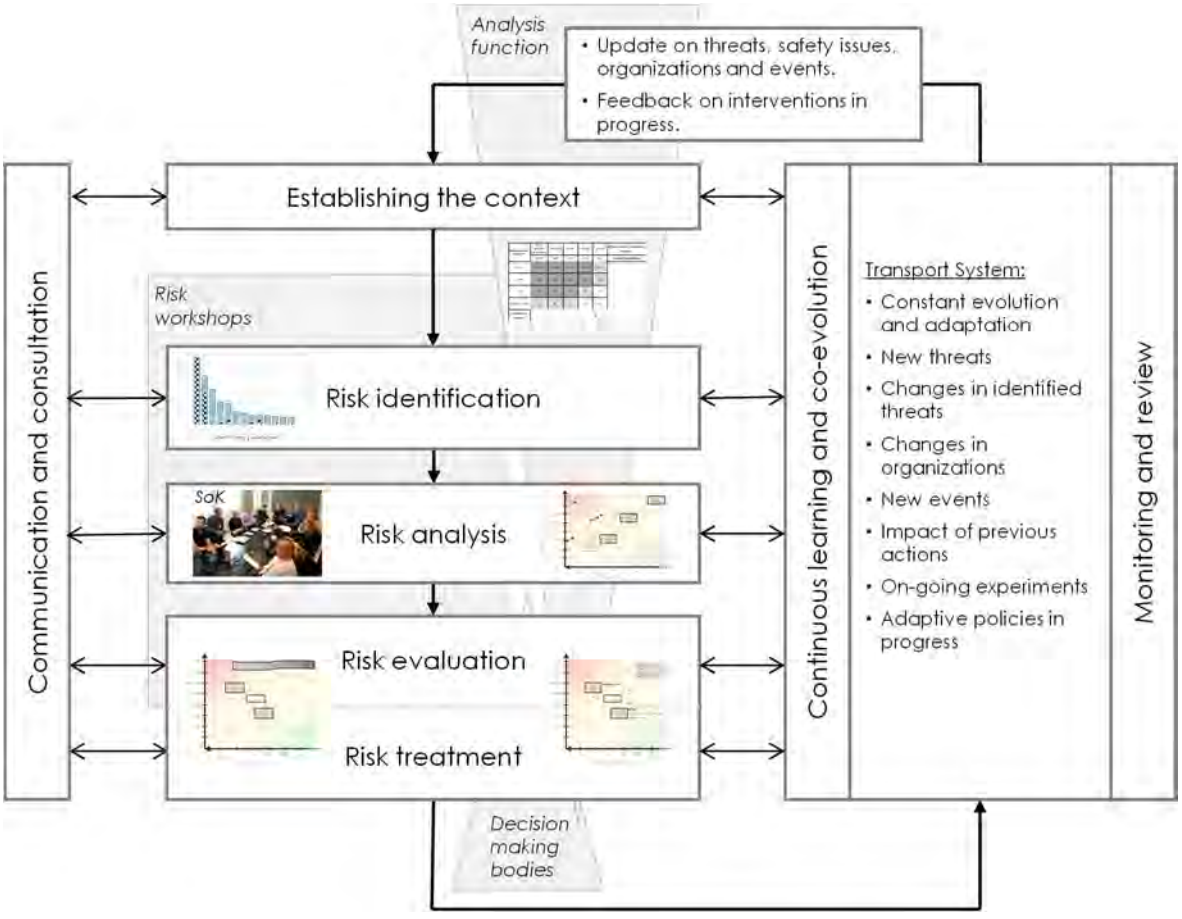


Figure 25. The proposed risk management process presented in the form of the ISO 31000 framework.

The organizational teams taking care of the various tasks within the process have been presented on the background. Symbols of various types of the risk picture, as well as some other items, have been presented at the different stages of the process. The event risk assessment matrix symbolizes the preliminary analysis tasks carried out by the analysis function. A bar chart at the risk identification stage symbolizes all the intermediate analysis results which can be used for risk identification and risk analysis. Picture of people attending a risk workshop symbolizes the important steps of integrating all

the data and information and building strength of knowledge (SoK). The transport system has been represented on the right side more explicitly than in the original ISO 31000 flowchart. The eight bullets within the transport system box remind of the different ways in which the system evolves continuously. The inputs to the risk assessment process, gathered from the transport system, are specifically highlighted in the two bullets on top of the flowchart. A new element has also been added next to the transport system to stress the importance of continuous learning and coevolution. This learning is linked with all the different stages of the process. Specifically, it is hoped that future decision-making improves thanks to constantly updated mental models of the system and its dynamics.

The communication and consultation activity on the left of the flowchart has to do with the various stakeholders. In the case of a national transport safety agency, the key stakeholders would include at least the political decision-making bodies above the agency, other state agencies, individuals involved in transport-related activities either professionally or through hobbies, and the public. In the interest of not cluttering the presentation, the stakeholders have not been presented one by one.

Overall, the proposed process is designed with deliberate compromises between the use of resource and the results in terms of risk assessment and treatment. Specifically, this is visible in the event risk assessment step, which could very easily become too heavy in terms of workload and lead time, if they used approach was not adapted to the volume of data. Such compromises are in line with the requirement (C-1). Being able to focus resources in risk treatment in the most effective way, in line with requirement (C-2), has been another leading guideline, featured through cross analysis between scenarios and organizations, as well as through adaptive approaches.

Many other requirements set for the risk management framework are met through the risk workshops. Knowledge of the people inside the organization is captured (RA-3) and diversity is created by inviting people from outside the core group and even from outside the organization (RA-4). The monitoring of organizations (RA-5) is done at least partly by the risk workshops, even if the systematic monitoring may be assured by some permanent body in the line organization. Proposals for action are made holistically and challenged from the perspectives of the non-safety priorities (RE-1). The workshops also provide an easy way to involve key stakeholders in the process (RT-7). The risk workshops and the decision-making bodies jointly ensure that decisions on risk acceptability benefit from the capability of people to consider different priorities, ethical principles and to make value judgments (RE-3).

7 Review: evolution of the process

No process will stand the test of time especially in a complex system. Some changes may be relatively easy to foresee, for example the introduction of new data channels. New technologies may transform the operational processes as well as make more data available. Some of the changes may be more surprising, like political decisions having an impact in the organizations carrying out the risk management. A massive reorganization with reallocation of responsibilities may be an example.

A proposed practice is to review the process at least yearly to see if adjustments are necessary or desirable for gaining improvements in terms of effectiveness and efficiency. In the case of a safety agency, this could be a shared task between the risk workshop participants, experts from the analysis function and senior management. Additionally, big changes in the organization or in the operational environment may require more immediate ad hoc changes.

There are some specific considerations related to data capture. In the case where not enough safety data is available, one easily becomes willing to get as much of any kind of safety data as possible. Different types of data give different perspectives to the operation all having their strengths and weaknesses in terms of the visibility they provide, but also in terms of the workload they introduce to the analysis function. Therefore, one should be cautious about what kind of data one wants to capture more, so that the information value per resource used remains acceptable. Sometimes a safety agency itself may have the opportunity to use its own resources in helping capture the data at the source. For example, if observational data from the bridges of ships or from the cockpits of aircraft could be made available

through a financial investment by the agency to hire suitable neutral observers, would this be a good investment? Observational data is very rich and useful because it provides the full context to what happened and how. Cooperation between operators and agencies around observational data could well be a better strategy for safety data capture than focusing on having more reported safety events.

8 Outlining a supporting software

There are several reasons why supporting the developed process with a dedicated software might be a good idea. This is not simply a question of moving work from individual excel files into a specific software but really also about discovering opportunities to carry out the different steps better, adding functionalities and introducing opportunities to discover things that would not otherwise be discovered.

There is a huge amount of data going through the process and higher-level information is created throughout the process. Results from previous stages are used as inputs for the later stages. All the data and information need to be kept safe and easily available to the people who need to work with them. With the number of items increasing, the 2D risk picture might become more and more cluttered and some new ways of presenting information dynamically may need to be developed. The amounts of data and the complexity of the transport system mean that there is a lot of value in offering an interactive way to analyze the data typically through visual analytics. These are probably the most obvious reasons why a software may produce significant value in supporting the risk management framework.

In addition, there are many secondary reasons. Simple tracking and alerting tasks can be conveniently left to the computer thus offloading the people. The software can keep track of all changes introduced to the data and all decisions made. It can produce automatic alerts and notices, for example to highlight a successful or a failing experiment. In the context where there is a lot of data and a lot of things may change constantly, updating things will be much faster with the software with dedicated functions than redoing individual charts. Overall, it should be expected that a proper software can help create better results but also save the related resources and time.

The following paragraphs outline the functions a dedicated software should carry out, thereby creating a first generic definition for the software. It is proposed, however, that the real development of the software would take place in interaction with the team who has already implemented or who is in the process of implementing the described risk management process - so that the dynamic interaction between the users and the developers reveals new opportunities and refines the functions and interfaces.

In many cases, the incoming data itself can be obtained in an electronic format. Even if this is not always the case, latest when the data reaches the first stages of processing it will be transformed into an electronic format. This means that the incoming data can immediately be hosted by the software. The software should feature an adapted interface for carrying out the first steps including categorization of safety events. The first significant step is the event risk assessment where the software can provide the necessary interface adapted to the specific needs, for example customized for each mode of transport. There are two main solutions for entering the result of the event risk assessment: picking a cell in the matrix or making a direct manual entry of specific severity and probability values. Experience from the case study (see Part III) shows that it is good to provide both alternatives, because depending on the case at hand, one or the other might be the most adapted solution. Such a flexible solution for event risk assessment is easy to implement with software. The user can either click on a customized matrix which produces the result directly, or use dedicated fields to enter specific values for severity and probability. The software can also accommodate the specific case where there are both actual and potential outcomes: first, the actual outcome is assessed with probability value 1 and the experienced severity of the actual outcome, and then the potential further escalation is addressed by carrying out an additional standard event risk assessment. The software sums the two results together and allocates this *riss* value to the event.

As soon as incoming data reaches the software, it becomes available for the whole community of users, to the extent that this is desired. It will be possible to see at what stage of the process the data is and for

example who has carried out the event risk assessment. Such information also helps calibrate different analysts for consistent assessment results.

In a similar way, the software provides the framework for safety issue risk assessment. It guides the analyst through the exercise step-by-step providing specific fields for the different entry points within the safety issue risk assessment method: exact definition of the safety issue, definition of the scenario and its different elements such as the undesired operational state, the barriers and finally numerical estimates for the probability and severity ranges. Strength of knowledge should also be captured so that the safety issue can be immediately visualized within the risk picture among other threats.

The software can offer the possibility to visualize elements in the risk picture in any chosen context. For example, after introducing a new safety issue in the risk picture the analyst might want to see all the other safety issues in the same mode of transport or maybe even their evolution in time as a live animation within the dynamic risk picture. One could also choose to plot all safety events related to similar safety issues or, again, where similar safety events are within the picture at different historical moments of time. This would show whether similar events get closer to an accident today than they used to some time ago. One can also examine safety issues and events which share some common safety factors. The opportunities are endless. An important part of the statistics are charts based on cumulative event risk values over time, between different locations, comparing different operators, and different types of operations. All such analysis work can produce statistics and interesting views to the data, at this early stage of the process - and feed the later stages of the process. It might also be useful if specific views and statistics could be stored and fed forward to a later stage in the process where the analyst thinks it could be appropriate to review those statistics.

The software can be a powerful tool for knowledge capture, not only for storing existing electronic data, but also for storing knowledge that people have – at different stages of the process. It should be possible to store information in the form of audio files and films so that recording the information is fast and easy for people willing to share the information. Ideally, the tool could take advantage of voice-to-text functionalities which again makes data capture easier. Obviously, the software should be able to store data files (e.g. audit findings) in a sensible way linking them with the relevant elements within the tool. It will be an ideal way to prepare risk workshops by making all the relevant material available to the participants in one place through the tool.

The two-dimensional risk picture is an important element of the process and becomes also an important part of the software. It should be easy to create new elements and to modify existing ones. Placing a new item in the picture could also be possible by dragging it to the right place in the picture. Like mentioned above, both data visualization and visual analytics should be available. The different views to the risk picture would not only include snapshots but also animations through time and zooming onto only a part of the big picture. As different elements typically have different exposure rates, it is very useful if the software keeps track of these and makes it easy to switch from type I to type II risk picture. It can also be envisaged that the software tracks some exposure rates automatically through the use of existing databases or thanks to the incoming data itself.

As clutter is one of the potential problems with the risk picture, the software can introduce several different presentation formats and symbols to ease this problem. For example, using crosses instead of rectangles to present threats takes less space in the risk picture.

The analytical functions related to the risk picture would include being able to visualize different entities together and compare them, drilling into any detail in the picture with the help of safety factors, categories or any other parameters. It is an advantage if results can be transposed on the same picture without the need to jump between different views. For example, a *mouse-over* a safety event can open up a caption with a short narrative of the event. In this way, several layers of information can be presented simultaneously helping to see both the big picture and some of the key detail simultaneously.

Drafting the potential interventions and their expected impact on the threats in the risk picture is challenging, not least because of the amount of information to be presented at the same time. Trying to visualize the interactions between the different threats and between the interventions is another difficult task where the software can be very useful. Part of the added value here is being able to present the visual information in an uncluttered way and to be able to choose only the wanted elements to be visualized. The software can support the definition of actions and experiments both at the level of the risk picture and at the level of defining the interventions in detail. The view at the level of the risk picture can be inspired by Figure 24 giving the experts an overview of different options and their cost dimensions. Each action and each experiment would then be defined in detail in a specified format and these definitions would be available for review and modifications directly through the risk picture. In addition to reaching the interventions through the risk picture, another important way to manage them is to present a timeline with all the ongoing and planned interventions. It is useful for the software to show how different experiments may have been combined together, stopped or expanded in different ways. It should be possible to overlay on the timeline also the current perception of success of failure of the different experiments and actions. The interface should support getting easily in touch with the right people responsible for the different interventions and reviewing historical communication threads with these people.

At each stage of the process, software logs all the decisions with their rationales. This includes work at the risk workshops and also specific decisions about the interventions. It should also be possible to open questions in the software that would remain visible and guide the work in the future stages of the process.

To reflect working on the three levels, the software needs to enable analysis and decision-making also at the level of organizations and at the level of the transport system. This includes visualizing the risks by operator and supporting cross-analysis where threats are mixed with operators to identify which combinations produce the highest risks. As difficult as it is to map interactions in a complex system and to try to understand how the system works, the software should support drafting parts of the transport system in the form of an interaction model which could be used as a reference and developed further when new experiences and ideas emerge.

Finally, as learning is a key objective of the whole process, this should be reflected explicitly in the software. It should be possible to log in comments and questions at each stage of the process and each level of the transport system. In other words, the software should support the human community working on the process not only in its everyday tasks but also on the long-term objective of constant improvement and learning. Lessons which emerge and gradually create an organically growing base of relevant knowledge become useful material for newcomers and other interested parties to learn about transport risks and their management in the current system.

9 Discussion

In many ways, a key challenge in developing the framework has been the fact that any model used for risk assessment would be highly simplistic compared to the extreme complexity of the real world. Event risk assessment and safety issue risk assessment are both carried out using a very simple and linear model relying on an estimate of probability and severity. The simplicity of the model becomes a real problem when an event does not directly link with an accident scenario. This is the case for example if a crew member becomes incapacitated: such a situation would likely increase the probability of many accident scenarios, and on the other hand make some specific scenarios related to additional failures particularly severe. The simple model allows to treat such cases only superficially without addressing the embedded complex logic. The simplicity of the models is easy to criticize, especially when a modern approach embracing complex adaptive systems is promoted throughout the work.

There are several arguments for using such a simple model: first, to process thousands of events, the initial part of risk analysis needs to be fast and simple, as in real organizations, the time and resource available are never unlimited. Cases such as the incapacitated crew member have such a complexity that it is difficult to think of any model good enough to replicate the dynamics of the system, except the real

system itself. Event risk assessment is only an initial estimate of risk and only an initial input to the risk management process, so rough estimates can be considered more acceptable, and the notion of low strength of knowledge can be associated with these results. The fact that events are *reality* and not only abstract threats, justifies trying to capture their contribution to the risk management framework, enhanced by the necessary expertise of the analysts. Later stages of the risk management process embrace more complex models all the way up to considering the transport system as a complex adaptive system.

The use of *safety factors* also introduces the danger for simplifying things too much. It must be stressed that the list of safety factors is not a universal complete set of factors guaranteeing safe operation. Furthermore, as mentioned above, there are nonlinear interactions between several safety factors. It is important to keep in mind the gap between the complex reality and a simple method like the safety factors. There is also not a single way to create a list of safety factors, which invites criticism that there may not be a repeatable scientific method for coming up with a set of safety factors applicable to a specific domain. Also, despite the attempt to define safety factors at the level of functions, the competing objectives of covering all potential safety-threatening factors and not having overlaps between the factors, seems to unavoidably result in safety factors which are at a more detailed level than a function. Furthermore, the line between the different levels is fuzzy. Despite these weaknesses, it can be argued that a carefully constructed set of safety factors which is tested in real use by domain experts, brings a new valuable perspective to operational safety - and more specifically provides valuable guidance on the risk reducing potential of different interventions, especially when cumulated risk values are allocated on all the safety factors. The importance is in understanding the strengths and limitations of the safety factors as an approach and using them in a consistent way while not making too simplistic conclusions based on the obtained results.

In using the safety factors, judgments will have to be made on which factors are seen as having failed or succeeded in a particular event. There will always be a boundary layer where the analyst could ask whether certain factors where *applicable enough* to be considered as failed or succeeded. There could be subjective differences in relation to this threshold: some analysts tend to become more safety factors than others. Especially on the side of positive safety factors, the guidance is to pick only the most important factors - often this could mean picking none. Otherwise, there would be too much noise in the results as most of the time the functions work successfully. Concerning all these challenges related to adopting the necessary mental models and getting used to applying the methods in the operational context, experience shows that questions will emerge throughout the learning period which typically lasts several weeks or months. Today's experience on the initial ERC method adopted by many airlines demonstrates that this is a fully workable method. Initial experience on safety factors is also very positive and will be discussed under the case study below.

Despite the simplicity of the probability-severity dimensions for event risk assessment, the mental work around assessing event risks often turns out to be challenging, especially for novices. The act of imagining an escalation into an accident may be quite difficult for banal events. If this mental escalation is compromised, then the assessment will produce too optimistic risk values. Some people also struggle with the concepts of *actual* and *potential* outcomes. There are similar challenges in safety issue risk assessment, in understanding what the existing barriers are, and in selecting the *undesired operational state* which acts as the divider between *avoidance* and *recovery* strategies.

The *risk picture* has a central role in the presented risk management framework. At first look, it could be perceived as a very banal and simple way to place risks inside a two-dimensional probability-severity space. The way it has been defined in the current work, however, produces a refined solution to many key challenges in risk management. It accommodates the hierarchy of threats, safety issues, scenarios and events. The important aspect of *strength of knowledge* can be made visible. The way the risks are assessed allows overlaying different dimensions of severity (e.g. environmental, material, human) on the same scale. The risk picture allows the comparison of different threats while taking into account their different exposure rates. The special role of black swans has been addressed and there is a dedicated way to handle such risks. The two-dimensional presentation allows to address the acceptability of risks

in a more adapted manner than a numerical treatment would. It also allows leaving the *risk aversion* policy open, i.e. it can be implemented through judgments based on the risk picture rather than as an inflexible constant value. All risks can be seen together and compared - within a single mode of transport but also across different modes - and decisions can be made based on the comprehensive picture as a whole, rather than on a case-by-case basis. The risk picture provides a good overview of the risk situation and produces a good basis for decision-making.

Only real-life use of the developed methodology will show to what extent the risk picture gets cluttered when all applicable data is presented. The cluttering is one of the reasons why it is probably wise to implement the risk picture through software. Therefore, some initial elements for such a software have been presented above.

The developed framework does not stay at the level of threats. A perspective to deal with risks at the level of organizations has been developed. Ultimately, the framework is positioned in the context of a complex adaptive system. The approach to decision-making and attempts to improve safety in the system are adapted to the complexity of the system. Focusing on the interactions between different risks and between different interventions is encouraged. In terms of the interventions with the system, experiments have a key role. This underlines the need for feedback and adaptation based on feedback rather than simple feedforward actions. The use of experiments and adaptive policies resonates with adaptive risk management.

Human knowledge, experience and interactive work done in risk workshops along with human judgment are considered backbones in the developed process. Continuous learning of the people involved is also seen as a key long-term asset for running a successful risk management activity. At the age of Big Data, such a reliance on people could be considered controversial, especially when the limitations of humans are taken into account in the context of risk assessment, decision making and ability to comprehend complex systems. Doesn't this endanger objectivity of risk assessment and make it very difficult to keep factual records of current knowledge and rationales for decisions? As a generic response, the underlying thinking here is that machines would only be able to take the decisions properly if the complexity of the real world could be programmed into the decision-making tool. This is not possible today, considering dimensions such as culture, motivation, discipline, dynamics of human error, various levels of compliance with rules and recommendations, etc.

Risk workshops could also be criticized for their resource requirements. It is argued that the resources are justified due to the key role of risk workshops in building the necessary knowledge for proper decision-making. To be successful, the risk workshops need to employ techniques which are in line with complexity: embracing diversity and coming up with several rationales and experiments rather than a single perspective.

10 Conclusions

A NRMF has been proposed. The framework has been created based on the detailed requirements developed in Part I. Therefore, it embraces modern concepts of risk and safety in the context of a complex adaptive system – in this case the transport system. The developed framework is the proposed answer to the research question.

In terms of key elements of the new framework, it allows combining different types of safety data, integrate them together and enhance the strength of knowledge as a base for risk assessment; it provides a holistic view to risks, giving the possibility to mix different modes of transport and different types of risks (e.g. human/material/environmental) together in a risk picture; it provides a solution for assessing the acceptability of risks and taking decisions when different strategic priorities along safety are all taken into account; it introduces a way to manage the so-called black swan risks within the framework and provides guidance for interventions with the surrounding system in a way which is adapted to a complex adaptive system; the framework provides a process for continuous learning and improvement with humans as the key agents.

The solutions proposed range from a new theoretical framework (e.g. the concept of riss) to specific presentation formats of the risk picture and to considerations on implementation of the methodology at a national transport safety agency and through specific software. The context of a national transport safety agency was chosen as the reference due to its superior level of challenges, e.g. compared to a single operator. However, the developed framework could be customized for another area of application.

The novel elements include a refined risk picture which can contain threats, scenarios, safety issues and even safety events; the concept of riss helping to manage both actual and potential consequences of past events in a risk assessment context; being able to mix transport modes and risk dimensions together, making all risks comparable on the same scales and proposing a full risk management process which can be used in the context of a complex adaptive system.

Due to the very challenging context shaped by complexity, fuzziness of human perception of risks and their acceptability, as well as the very low probabilities of many transport risks leading to lack of hard data, the proposed framework is not a detailed recipe or a mathematical solution: it leaves room for human judgment which is necessary for dealing with the inherent complexity and the multi-dimensional value considerations.

PART III – Validation

Chapter 8 - Validation of the proposed risk management framework

Ideally, the developed method would be validated by presenting hard data on its performance. There are two reasons which make such an approach impossible: the scarcity of accident data in a very safe system combined with the intractability of a complex system. The only acceptable data for proving that risks have decreased is accident data. The use of incident data is problematic because of the questionable link between incidents and accidents, as discussed in Chapter 4.1.1. As a big part of the transport system is very safe or even ultra-safe (see Chapter 4.2.2), data is not frequent enough to prove anything. It could take years to show a statistically significant improvement, even when assuming a completely stable system: Amalberti (1999) quotes a delay of four years for a system with a 10^{-5} accident rate and about six years with a 10^{-7} accident rate. Moreover, the system is anything but stable. As a complex adaptive system, it is in constant evolution and there are thousands of changes and change agents acting all the time. It would be impossible to attribute a specific change in safety performance to a specific change in the system, for example to a new risk management process. As stated in Chapter 3.5.3, it is simply impossible to conduct controlled experiments in complex systems. Even if the accident data was more frequently available, the problem of determining what are the causes for having more or less accidents would still remain. This is particularly true in a transport system where accidents are no longer related to simple technical failures but more associated with emergent organizational characteristics such as safety (culture) and resilience.

Implementing the NRMF in different organizations could produce valuable feedback on various aspects. For example, if the sub-methods were too difficult for the analysts or if the workload introduced was way too excessive, this would strongly suggest that the approach is difficult to implement in practice. However, success or failure in implementing the NRMF would tell at least as much about the organization itself than it would tell about the framework. What are the safety management paradigms in the organization before the implementation and how well do they support the new paradigms coming with the new framework? Does the new framework fit with the existing strategy and priorities of the organization? How much does the top management want the implementation and push for it? How much resource is there in the organization and how much of it is allocated to making the implementation work? What are the knowledge and skill levels of the people involved? How much external resources are used to help in the implementation? Such questions are very relevant but it is very difficult to assess such factors. Additionally, when the implementation starts, a co-evolution starts between the change agents and the rest of the organization. The two sides influence each other. The organization may change beyond the strict scope of the framework itself, and on the other hand the framework will undoubtedly be customized to better fit the existing characteristics of the organization. So *once the implementation has started, the organization is no longer what it used to be, and the framework is no longer what it used to be*. The case study below provides examples of such phenomena. Another practical constraint for validation using several organizations is that there are not many multi-modal transport safety agencies in the world.

It is concluded that any simple *quantitative* approach of validation would mean closing eyes to these facts and be illusionary. The adopted validation approach is based on three components which access the proposed framework more in depth than quantitative approaches.

First, the developed framework is compared with the established requirements for the framework. The requirements were carefully constructed based on the literature review reflecting the desired modern approaches to risk management and to interventions with complex systems. Showing that the framework complies with all the requirements is thus a key part of the validation. Secondly, the framework is compared with existing frameworks to illustrate the added value. The developed requirements are used to support this comparison. Finally, applicability is demonstrated through a case study covering the

customization and partial implementation of the approach in one multi-modal national transport safety agency. In describing the case study, special attention is paid to treating the safety agency and its environment as a complex adaptive system and trying to highlight the influencing factors and rationales behind the observable behaviors.

1 Validation against the developed requirements

Reference to the requirements was already made in Part II when the developed risk assessment framework was introduced. A summary of how the framework meets the requirements is presented here. Each original requirement is presented first as a bullet and the way the requirement is complied with is presented as a sub-bullet(s).

Risk Identification

- (RI-1) As the transport system is constantly evolving, and so are the risks within, the RMF needs to include provisions to feed in a continual flow of data/information from the system.
    - The cyclic process produces continual updates on threats, safety issues, organizations and events.
- (RI-2) There is already an established flow of *safety information* within the transport system. The RMF needs to be able to digest *the types of safety information* used currently and transform such data into useful risk information.
    - Current safety information includes event related information in the form of categorized narratives, safety indicators and statistics, results of studies, oversight-related findings and knowledge, safety-related information on organizations.
    - The framework can capture virtually any type of safety data/information: event data can be transformed into risk information through event risk assessment; other types of existing information are captured in the knowledge-building at the risk workshops in the form of intermediate analyses, statistics, as well as relevant information and knowledge. They all contribute to identification of risks and their analysis.
    - Events detected in flight data would not enter as such. They would need to be processed through event risk assessment (but this step could be automated). Another option is to enter safety issues detected from flight data directly into the risk picture as issues.
- (RI-3) The need to carry out a risk assessment related to a safety issue or a change must be one of the normal inputs to the process.
    - Risk assessing a safety issue (including a change) and placing it in the risk picture has been specifically addressed.
    - In some cases the impact of a change can be addressed also at the organizational level, impacting the resilience.

Risk Analysis

- (RA-1) The new risk perspective (incl. focus on uncertainty and surprises) should be embraced.
    - The new risk perspective is used as the paradigm for risk analysis throughout the work.
    - There is a specific knowledge-building step in the process. Uncertainty is explicitly presented in the risk picture. Surprises are addressed through the special treatment of the risks in the black swan-corner.
- (RA-2) There needs to be a specific focus on *building knowledge* prior to risk evaluations and decision-making.
    - There is a specific knowledge-building step in the process, taking place at the risk workshops.
- (RA-3) Knowledge of the *people within the organization* (e.g. the safety agency) needs to be captured, including tacit knowledge.
    - This is addressed through the knowledge-building in the risk workshops and reinforced by the recommended ways to run the workshops (counteracting known biases). Information is not passed to another team without knowledgeable people coming with the information.

- (RA-4) *People from outside the core group* (in the agency) and *operational people* from the field should be pulled into the process (to capture knowledge and to create diversity).
    - Both groups of people have been specified as participants to the risk workshops.
- (RA-5) One should aim at understanding the *resilience* of the organizations carrying the risks, and to find ways to describe it, preferably in a comparable way.
    - The organizational level has been specifically addressed, including monitoring of key organizations and their safety/resilience characteristics.
    - An initial set of organizational safety factors have been proposed. These could be used to track organizational resilience. The main challenge is to gain access to organizations at the depth required and the related effort.
- (RA-6) The RMF should feature a *holistic risk picture* including risks/threats from the 4 modes of transport and allow comparisons of risks *relative* to each other.
    - The risk picture is a core part of the proposed framework. The type II picture allows comparisons between different risks.

Risk Evaluation

- (RE-1) The decision process should be such, that the decision on the acceptability/priority of a risk or a solution can be done in a *holistic manner*, *taking into account the assumptions* and the *strength of knowledge*, while paying attention both on the risk and on the *costs of its treatment* and considering the *various priorities, not only safety*.
    - All the decision elements are available: the background information (including assumptions) from the knowledge-building stage and the uncertainty indicated in the risk picture.
    - Presenting the intervention options in the risk picture provides the opportunity for making decisions where both the risk criticality (position in the picture) and the treatment costs of various intervention options can be considered holistically.
    - Challenging the intervention options from the point of view of the various priorities is a specific step in the process.
    - Adaptive approaches (experiments and adaptive policies) spread the decisions over time and enable more factual information (e.g. experiment results) to be used in the decision making.
- (RE-2) Consequently, ideally the risk picture should be able to host non-safety risks (e.g. financial, reputational risks) in a compatible manner.
    - Due to the generic nature of the scales in the risk picture, it should be able to host virtually any kind of risks.
- (RE-3) Decisions on risk acceptability need to be done by humans, due to their capability to consider ethical principles, various conflicting priorities and make value judgments; i.e. the risk management methodology does not need to (and should not) try to produce the final answers directly.
    - The process does not provide answers automatically, but rather tries to provide the best possible circumstances for decision-making by the best-suited people.
- (RE-4) The *black swans* need to be addressed and the high-impact–low-probability scenarios must not be dismissed solely due to their low probability.
    - Risks in the black swan-corner (top-right) have a special treatment and they are highlighted in the risk picture in a way which makes them visible independent of their theoretical probability values.
- (RE-5) Ideally, the risk *aversion policy should be left flexible* (i.e. not pre-determined).
    - The risk picture presentation keeps the severity and probability dimensions separate and does not propose pre-defined risk acceptance criteria. Acceptance is done through a study of the overall situation, taking advantage also of the other risks in the picture. This leaves the risk aversion policy (in terms of big vs. small accidents) open.
- (RE-6) It should be possible to apply specific industry references (e.g. such as by IMO, RSSB or EASA) related to risk acceptance or severity assessment (e.g. “1 fatality corresponds to 10 severe injuries”).

  - Specific references can be easily implemented thanks to the system of severity points. With an intelligent software implementation, sensitivity to such adjustments can be studied even afterwards: e.g. "how would the risk picture change if 20% more severity points in aviation accidents were given to people on the ground, compared to pilots?"

Risk Treatment

- (RT-1) *Risk treatment options* should also be presentable in the risk picture with some measure of their associated "costs" (in a large sense), so that *alternatives can be compared*.
  - A version of the risk picture including treatment options has been presented. Using a specific software can provide a lot of opportunities to optimize the presentation format.
- (RT-2) There needs to be a way to *present and properly address black-swan-risks* and other very low probability risks, e.g. by building suitable *resilience* in the system.
  - Specific (resilience-based) ways to treat these risks have been proposed.
- (RT-3) The interventions with the system need to be *adapted to the nature of the system* (or sub-system), e.g. referring to Cynefin.
  - Use of the Cynefin framework has been introduced as a specific step in the process for highlighting the nature of the system/problem at hand.
- (RT-4) As the transport system is seen as *a system of sociotechnical systems*, it is important that interventions embrace the typical features of CAS, including non-linearity, unpredictability, counter-intuitiveness, unintended consequences of interventions and the fact that emergent system properties like safety cannot be controlled directly.
  - The CAS-nature of the transport system has been stressed throughout the development of the framework. Adaptive approaches, such as experiments and adaptive policies, have been promoted as the most promising interventions for complex systems.
- (RT-5) An *adaptive approach* needs to be adopted for risk management and interventions. A recommended *modus operandi* is to use parallel experiments (e.g. small-scale implementations) and to adapt them based on the feedback e.g. by expanding the successful interventions.
  - Adaptive approaches, such as experiments and adaptive policies, have been promoted as the most promising interventions for complex systems, and managing the portfolio of such on-going interventions has been discussed.
- (RT-6) A holistic view of the whole must take precedence over fragmented approaches because in a CAS, different risks (and interventions) are typically interconnected (Ackoff's "*mess*").
  - The idea of reviewing the risks and the interventions as a whole and trying to map possible interactions is part of the process. Adaptive approaches like experiments reveal interactions between interventions so that they can be addressed.
- (RT-7) *Stakeholders* need to be involved systematically to give their inputs and feedback.
  - Involving the stakeholders e.g. through the risk workshops has been included in the process.
- (RT-8) While a complex system can never be fully understood/described, constant learning *about the transport system, its risks and its reactions to different types of interventions* should be facilitated. Such *coevolution* between the decision makers and the system can be seen as one of the key objectives of the process and improve future interventions.
  - The continuous learning and gaining wisdom have been specifically discussed as a part of the framework. The existence of the various teams around the risk management activity – and especially the risk workshops – provide the groups of people where the learning takes place.
- (RT-9) Irreplaceable human characteristics (e.g. sensitivity to history, context, objectives, values) should be exploited in making sense of the CAS, but distinct features in the process should counteract human biases and limitations (e.g. availability heuristic, groupthink).
  - In the context of learning and coevolution, a list of strategies for counteracting the limitations at least partly has been provided. These strategies can be embedded in the risk management process.
- (RT-10) Both from the intervention and learning points of view, three distinct levels need to be considered: events (system *behavior*), organizations (elements) and the system level (parts of,

or the whole transport system); and besides *analysis*, there should be an effort to apply *synthesis*, i.e. study the role the system plays within the larger context, and the objectives/constraints from above.
    - All three levels have been discussed in the context of the methodology, and the way these levels can be addressed has also been pointed out in practice. For example, monitoring the organizational safety factors could be carried out by the oversight function or the risk workshops. Applying synthesis to better understand the phenomena under study has also be specifically included in the process.

Context

- (C-1) Safety and risk management functions need to be carried out within acceptable levels of *efficiency, resource and cost*. This is particularly critical for process steps which are performed frequently (e.g. event risk assessment).
    - Maintaining acceptable levels of pragmatism and workload has been a key concern in designing the process. Constant building and re-building of detailed system models and detailed analysis of every incoming safety events are ruled out. The repetitive activity of event risk assessment is made limited in its scope of analysis and the resulting low strength of knowledge is an accepted compromise. Precious time in risk workshops is saved thanks to preparatory steps being carried out by the analysis function.
- (C-2) Thanks to the process, the Agency should be able to focus its resources *as precisely as possible on the most critical risks* (which could be combinations of risk scenarios and organizations, for example).
    - The risk picture shows the criticality of the risks. Targeting the interactions as precisely as possible is supported by the link to actual events and their safety factors (e.g. specific competencies) and the cross-analysis combining scenarios and organizations. Thanks to event risk values, safety factors can be compared based on their associated risk.
- (C-3) In addition to inputs from Risk Identification, it should be possible to integrate into the process *proposals coming from outside*, e.g. from the political system, and make sure such items are comparable with other elements.
    - The process has a provision for hosting proposals within the risk pictures, both as risk elements and proposed interventions. Once in the risk pictures, these elements are fully comparable with the other elements.

This summary shows that each requirement has been specifically addressed in the NRMF. Therefore, it is a modern framework, embracing the new risk perspective and the complexity of the transport system, while accepting the pragmatic constraints related to real-world implementation.

The challenges of complex systems can never be fully compensated even with the best of methods. However, it is argued that there is a dramatic difference between existing methods which do not address complexity consciously in any way, and the NRMF which identifies complexity as one of the key challenges from the beginning and implements a long list of techniques for that purpose. The Cynefin framework is used to identify the type of system one is dealing with and several techniques adapted for complexity are used, like adaptive policies and experiments. The need to focus on the whole at the system level and the interactions between problems and between solutions are emphasized. Involving a diverse group of stakeholders is encouraged, helping in creating many different perspectives and in collecting knowledge. The proposed process includes the cycle of constant learning on the system and the success of the interactions. With all these conscious efforts, it can be expected that the NRMF handles complexity significantly better than existing methods.

The main weakness is that the process needs to be implemented in an organization. This makes the NRMF vulnerable in terms of how well the people involved understand the various components and how well they are implemented. Availability of resource will be a critical success factor. Additionally, even if the initial implementation is a success, in the pace of time, key people and key knowledge may be lost, and there may be drift in the practical implementation. For example, the knowledge-building step in the risk workshops might get less and less time and attention, or a dominating person might get too much influence over the collective judgments. As an organization is a living organism, it is

impossible to guarantee that a process implemented by the organization remains adequate over time. Another potential weakness is that the risk picture may become cluttered with too many elements. Fairly sophisticated software solutions may be necessary to overcome this problem. This potential problem is hard to assess, as this type of risk picture has not been used before.

2 Comparison with existing frameworks

The next validation step is to examine whether the new framework produces tangible advantages compared to existing risk management frameworks. For this purpose, 20 different methods related to risk management were reviewed. The aim was to find methods which can be used in an operational context and which could fulfill as many of the requirements as possible. Both methods from the industry and from the scientific literature were reviewed. The most pertinent methods are presented in Table 1. These are the same methods which were reviewed in Chapter 6. The other methods are discussed shortly at the end of this section.

Table 1 presents a comparison of eight methods, based on the 28 requirements developed in Part I of this dissertation. Due to space constraints, the requirements are not repeated in the table in full, but another version of the table can be consulted in Appendix 4 where the requirements are reproduced in full length. The first five methods are industry methods and the following two from scientific literature. The NRMF is presented in the last column. Clear compliance (Y) or non-compliance (N) with a requirement is indicated with a single letter. A difference is made between non-compliance and non-applicability (N/A), for example for risk treatment -related requirements when a method only addressed risk assessment but not risk treatment. If compliance with a specific requirement is not straightforward enough, a short textual explanation is provided. This would be the case for partial compliance, or when the required feature has not been specifically designed in the method or reported but could at least in theory be implemented as a part of the method. The latter case would often by indicated with the comment “not explicitly covered”.

It is acknowledged that the assessments in the table are subjective assessments made by the author. It is argued that having the established set of specific assessment criteria should support producing consistent results – especially when any cases that are not straightforward are covered with a textual comment. The assessment information necessary for filling the table come from the key sources mentioned in Chapter 6, where the methods were first introduced.

The first observation is that none of the seven existing methods fulfill all the requirements. In fact, surprisingly few requirements are properly addressed. The set of requirements can only be fulfilled by a method which demonstrates both breadth and depth: the whole process needs to be covered including all specifics – but each stage in the process also needs to be covered properly, embracing the modern approaches. Some methods fail mainly in breadth and some others mainly in depth. The ICAO method covers almost the whole process, but very superficially without addressing any of the real challenges. The FSA method of IMO is not really a continuous operational risk management method but rather a one-time assessment, and it therefore lacks the breadth, but features some positive characteristics such as highlighting the interdependence between various risk controls.

The RSSB method is essentially a large safety model of the railway system which helps estimate risks within the system, without any risk treatment considerations. But it is a continuous operational method, in use, and fed with real operational safety data periodically. Unfortunately, the risk perspective is very classical, so both the risk assessment and risk treatment requirements remain unmet. The CASA method covers the whole breadth and features several positive elements, benefiting both from the operational context and the rather up-to-date ISO-based framework. It has its weaknesses mainly in depth of the approach for risk analysis and evaluation, and the lack of embracing the complexity through adaptive risk treatment techniques. ARMS is sensitive to operational needs but features a classical risk perspective and does not cover the risk treatment part at all. The framework by Aven implements the modern risk perspective perfectly and is clearly the most advanced from the existing methods in this sense. Because it has not been customized to an operation with a continuous, rich data flow, it does not

specifically feature a way to input event data. The Leveson method is an interestingly different approach which makes key safety assumptions very explicit. However, it tries to manage risks without assessing them and is based on a rigid model of assumptions and indicators. Neither modern risk perspective nor complexity is embraced.

Requirements	ICAO	IMO	RSSB	CASA	ARMS	Aven 2014	Leveson 2015	Nisula 2017
	Safety Management Manual (SMM)	*Formal Safety Assessment (FSA)*	*Safety Risk Model (SRM)*	*Regulatory Safety Management Program*	*Event Risk Classification and Safety Issue Risk Assessment*	*Methods from "Risk, surprises and black swans"*	*Risk management through leading safety indicators*	*NRMF*
RI-1	Y	N/A	Y	Y	Y	No explicit entry points for operational data.	Yes, but only related to the indicators.	Y
RI-2	Yes in principle. No method to transform into risk information.	N	Yes, at least the typical incident data.	Y	Yes for events, and other flows could be adapted.	No explicit entry points for operational data.	N	Y
RI-3	Yes in principle.	N/A	Not part of the basic process. Could be an ad hoc study.	Safety Cases can processed as an input, but the comparison is not visual.	N/A. No risk picture.	Y	N/A	Y
RA-1	N	N	N	N	N	Y	N	Y
RA-2	N	Not explicitly. Importance of data is stressed, ref. "confidence" in results.	No. Based on fixed model, which is assumed valid.	Yes, using several data sources and various meetings.	N	Y	N/A	Y
RA-3	Not explicitly covered.	N/A	No. Limited to building the model.	Y	Not explicitly covered.	Y	Yes, during the definition phase.	Y
RA-4	N	N	No. Based on a fixed model. Data from operations.	Yes. Involvement of stakeholders.	N	Yes. "Experts outside the core group"	N	Y
RA-5	N	N	N	N	N	Not explicitly. But importance of resilience is recognized.	N	Y
RA-6	N	No. FN-diagram could be used for this, but this is not part of FSA.	Partly. Single picture for railway risks.	N	N	Yes. The examples suggest how this could be done.	N	Y
RE-1	N	Partly. Only cost-benefit assessments are considered.	N	N	N	Y	Partly. Assumptions are explicit.	Y
RE-2	N	N	N	N	N	Y	N	Y
RE-3	No. Answer seems to come directly from method.	No. Emphasis on data.	Yes. Method gives risk levels only.	Not specified. Decisions seem to be taken in meetings.	Partly. SIRA gives acceptability directly, ERC is more flexible.	Y	No. Actions are pre-defined conditional to indicator values.	Y
RE-4	N	N	No. Low p scenarios add up to the total risk value, but are not treated as potential black swans.	N	N	Y	Not directly, but a black swan could show that an assumption was wrong.	Y
RE-5	N	Y	Y	N/A	Yes. Depends on customization.	Yes, implicitly.	N/A	Y
RE-6	N	Y	Y	Yes, implicitly.	Yes, implicitly.	Not explicitly covered.	N/A	Y
RT-1	N	Partly, in the form of cost-benefit assessments.	N	N	N	Not explicitly covered.	N	Y
RT-2	N	N	N	N	N	Y	N	Y
RT-3	N	N/A	N/A	N	N/A	Not explicitly covered.	N	Y
RT-4	N	N/A	N	No. But stakeholders consulted widely.	N/A	Not explicitly covered, but adaptive approaches introduced.	No. But a large part of the larger system is considered.	Y
RT-5	N	N/A	N	N	N/A	Yes, adapted approach introduced in principle.	N	Y
RT-6	N	Partly. Only risk control interdependencies are mentioned.	N/A	N	N/A	Not explicitly covered.	N	Y
RT-7	N	Not explicitly covered.	N/A	Y	N/A	Yes, indirectly, e.g. discursive strategy.	N	Y
RT-8	N	N	N/A	Not explicitly covered.	N/A	Not explicitly covered.	Partly, through the indicators and assumptions.	Y
RT-9	N	No. Strive for "objectivity".	N	N	N/A	N/A	Yes, biases explicitly addressed.	Y
RT-10	N	N/A	N/A	Yes, at least these levels are explicitly recognized.	N/A	N/A	Partly. The system level may be incomplete.	Y
C-1	Y	N/A	Partly, in its limited scope of the model. But updates are rare.	Yes, implicitly.	Yes. Workload and speed are key concerns.	Yes, resource limitation is recognized.	No. Difficult to estimate resource requirements, but probably very high.	Y
C-2	N/A	N/A	Yes, provided results are correct.	Yes. At least should be possible.	Yes, provided results are correct.	Not explicitly covered.	No. The approach does not deliver priorities.	Y
C-3	N/A	N/A	N	No, no global picture where comparisons could be made.	N	Y	N	Y

Table 1. Review of properties of existing risk management frameworks.

As presented in the previous section it is argued that the developed NRMF meets all the requirements at an acceptable level. It is argued that the proposed framework covers the full breadth required from the risk management process, it is adapted to the operational context, it embraces the modern risk perspectives and addresses the complexity of the transport system by applying adaptive risk treatment approaches. Therefore, a gap is observed between the new framework and the existing methods, at least at this stage when the NRMF is treated as a theoretical framework, assuming that it is implementable.

The rest of the reviewed methods were not included in the comparison table because they were considered less compliant with the requirements than the seven presented methods. The following industry methods were reviewed but not included:

- European Railway Agency (ERA) guidance for the *common safety method* (ERA 2009). The flowchart on page 26 and the statement [G3] on page 37 reveal that the method is made for being used for assessing the safety impact (risk) introduced by "significant changes" in the system. Implicitly, for the rest of the time, the system is assumed "safe" and the method does not therefore lend itself to operational risk management of "permanent, everyday risks".
- *Mission risk assessment methods* (often in use in the military) such as described by Johnson (2007, table 1). These tools are excellent demonstrations of how risk assessment can be adapted to be pragmatic and fast enough even for live military operations. However, they need to be designed for a specific domain and could not thus be used to handle data flows of transport risks with changing topics, nor to manage risks over time.
- *Organization-level risk assessment tools* in use at some state agencies to classify operators in different risk levels. Such methods are only one (potentially valuable) component in the overall risk management framework. Agencies use the results typically to define the frequency of audits in the operators. See for example the method of the Swedish Transport Agency where the operators score points based on defined criteria (Transportstyrelsen 2014).
- The *handbook of incident and accident reporting* cites the five stages of risk management in the US Army: identify and assess hazards, develop controls and make risk decisions, implement controls, supervise and evaluate (Johnson 2003, pp. 590-592). The techniques associated with the described method do not add new elements compared to the already reviewed methods.
- The *Bowtie method* which is essentially a static (simplified) model of one threat or scenario, based on barrier-thinking, and as such does not feature a real risk *management* process. See e.g. http://www.caa.co.uk/Safety-Initiatives-and-Resources/Working-with-industry/Bowtie/

The following methods presented in scientific literature were reviewed but not included:

- *Ship collision risk analyses* based on Bayesian belief network model (see Hänninen & Kujala 2012, Hänninen et al. 2013). The used model needs to be hand-made for a specific accident type and the method cannot thus be considered a comprehensive operational risk management method capable of dealing with all kinds of threats and scenarios – not to mention different modes of transport.
- *Multi-agent dynamic risk modelling* (Stroeve et al. 2013). The approach is applied only on one accident scenario, so the scope is not at the total system. Furthermore, the approach is based on a complicated model (with associated assumptions), which becomes potentially too limited in the context of a CAS which is organically changing continuously.
- *The functional resonance analysis method FRAM* (Hollnagel 2004, pp. 173-176, 2012, Bjerga et al. 2016). As the name indicates, FRAM it is really a model rather than a process for risk management. While such methods can certainly provide valuable insights, the effort of building all the necessary models to cover transport operations is likely too intensive for such an approach to be used in an operational setting.
- *The framework for risk management decisions in aviation safety at state level* (Insua et al., in press). The described method is fully based on mathematical modeling using historical safety data as a source. Such an approach is seen incompatible with the notions of complex adaptive systems and the modern risk perspectives which embrace uncertainty and surprises.

- *Integrated method for risk management in transport "trans-risk"* (Jamroz et al. 2010). The stated intent is to develop an integrated method of risk management across all modes of transport. The method as reflected in the paper is not yet in a detailed enough form to be considered.
- *A quantitative model for aviation safety risk assessment* (Shyur 2008). The method is another modelling approach and fed entirely by historical data. Again, such an approach is not dynamic enough to be used as a continuous risk management method in an operational context.

3 Utility of developed concepts

Besides the delivered NRMF, the related conceptual development adds to the existing body of risk research. The most important new concept is probably the concept of *riss*, which was discussed in Chapter 7.2.2. This new concept supports risk identification and risk analysis by enabling to combine data on materialized and non-materialized risks (i.e. losses and risks) and to create cumulated estimates of severity and risk. Without this concept, risks and losses could only be assessed separately which would not give a good overall measure. This was illustrated through an example in Chapter 7.2.2. It is intuitively obvious that if a certain issue (e.g. a new traffic arrangement) has produced both accidents and reported dangerous events, both types of experiences add to the perceived risk related to the issue. The use of the riss concept and the severity (and risk) points allows to assess both types of events on the same scale and to addition their impact. The cumulative values then enable comparisons between different issues in terms of severity or risk.

The *risk picture* in its proposed format transforms several concepts into a visual format: e.g. threats, scenarios, events, uncertainty and high-impact risks for which the assumed probability value becomes irrelevant. From the conceptual point of view, perhaps the most interesting feature is *the possibility to combine (current or future) risks in the same picture with past events*. Thanks to the event risk concept, the position of the events illustrates "how close" to an accident the event propagated, and how severe the accident would have been (if it had materialized). Even if the event risk part is based on a judgment, the events themselves are factual and positioning the events in the picture gives a complementary reality check to the "theoretical" threats which are linked to these events. For example, if all events are *higher* in the risk picture than the underlying risk rectangle itself, the reality seems to suggest that the *severity* level of the risk is higher than the current value reflected by the rectangle.

The *safety factors* are an interesting implementation of the Safety-II thinking and as illustrated in the next chapter, they are useful both in making sense of event data and as a framework for identifying operational resilience.

The *event risk assessment* creates a new currency for any type of statistics based on event data. Instead of using the event *count* as the currency – like in safety indicators – the currency is now the cumulated (event) risk. This is a move from a single-dimension to two dimensions, including the severity. Using fatalities as the main driver for severity and the severity point system (bringing injuries etc. on the same scale) is a much more justifiable basis for severity estimates than the original ARMS ERC severity classes and corresponding risk indices.

Chapter 9 - Case study: The Finnish Transport Safety Agency

The author cooperated during two years with the Finnish transport safety agency, Trafi. The starting point for the cooperation was the shared interest in developing a risk management framework for such an agency, so that the *modus operandi* of the agency could become more risk-guided. For this purpose a project was set up at Trafi. The project was given the name TiTo, coming from the Finnish words *Tiedosta Toimenpiteisiin*, i.e. *from data to interventions* - stressing the point that the new risk management framework would cover the whole process from the very beginning to the very end. The project ran from April 2013 to March 2015. Many of the components for the framework were developed during the project. The implementation at Trafi started gradually during the project but some of the major parts have been implemented after the project.

Besides the expressed interest of Trafi to support the development and to implement such a process, the fact that Trafi is a multimodal agency made it a very interesting and challenging place to implement a new modern risk management framework. Trafi was created in 2010 by combining several existing safety agencies and covers aviation, maritime, railway and road safety. This is a rare set up even at an international scale. From the point of view of risk management this means that the methodology has to be able to deal with four different modes - each of them bringing along a different world of traditions, practices, mental models, beliefs and terminology. As discussed earlier, besides the four modes of transport, the agency needs to manage its relationships with many stakeholders including the ministry of transport and the public.

The case study is split into three sections. The first section presents the starting point of the project, i.e. what Trafi was like at the beginning of the project. The second section explains the developments carried out during the project. The third section concentrates on the results of the project, i.e. what have been the developments at Trafi up to the writing of this dissertation.

1 Trafi at the beginning of the project

At the beginning of the research project (April to September 2013) the author carried out a review of Trafi, focusing above all on the safety/risk management aspects. The status of the agency in 2013 was the starting point for developing a new better risk management framework and provided the most important contextual aspect for the development. This first analysis was based on interviews and observation of current working methods and documents. The next two sections summarize the key findings.

1.1 Trafi mission, strategic objectives and organization

At the time of the project, Trafi defined its activity with the following statements (Trafi 2015b):

> *Trafi develops the safety of the transport system, promotes environmentally friendly transport solutions and is responsible for transport system regulatory duties.*
>
> *Trafi:*
> - *issues permits, regulations, approvals and decisions and prepares legal rules regarding the transport sector;*
> - *arranges examinations, handles transport sector taxation and registration, and provides reliable information services;*
> - *oversees the transport market as well as compliance with rules and regulations governing the transport system;*

- *participates in international co-operation;*
- *ensures the functionality of the transport system even in emergency conditions and when normal operations are disrupted;*
- *creates opportunities for innovative development of intelligent transport;*
- *informs the public of transport-related choices.*

Vision:
Responsible traffic

Mission statement:
We enable well-being and competitiveness from transport

Values:
Courage and co-operation

Strategic goals:
- *Influencing: Trafi shows the way and actively influences the drafting of transport policy and fulfilment of transport policy objectives.*
- *Customers and services: Trafi is a pioneer in customer-oriented public services.*
- *Information: Transport system development and provision of services to traffic is based on active utilization of information.*

More than 500 people work at Trafi. From the organizational point of view, it is challenging to match together a set of necessary functions for the agency together with four modes of transport. During its existence, Trafi has had several reorganizations. At the beginning of the research project, the three main functions of Trafi, also reflected in its organization, were:
- Regulation
- Permits, authorizations, licensing
- Oversight

In addition to these functional units, there were four nominated directors - one for each mode of transport.

There was a clear recognition in Trafi that the transport system has to be developed as a whole and that safety cannot be addressed in a vacuum without considering other strategic objectives (Trafi 2012, p7). Trafi had identified four strategic objectives going far beyond the safety mission (Trafi 2014a, p8):
- Safety and security
- Sustainability (environmental, social and economic)
- Functionality and reliability (infrastructure and services)
- Development of the commercial transport market

These objectives can also be seen embedded in the Trafi activity definition above. However, as reflected in the full name of Trafi, the safety mission was still the number one *raison d'être* for the agency.

Another dimension of Trafi's work is its role in relation to the European safety agencies and the international organizations dealing with transport safety. European safety agencies - such as EASA (European Aviation Safety Agency) and EMSA (European Maritime Safety Agency) - are actively working on safety regulation and oversight and this introduces a considerable amount of work for national agencies such as Trafi. There is work related to contributing to the development of new regulation and standards and reacting to proposals from the European agencies and to varying extent participation in the governance of these agencies. There is similar work related to the international organizations - such as IMO (International Maritime Organization) and ICAO (International Civil Aviation Organization). Every mode of transport is different in terms of what kind of decisions are made at the international, European and national levels, but binding legislation/regulation must be

implemented nationally, and the necessary resources must be allocated for this. For the management of Trafi there is a strategic decision to be made on how much resource is used to contribute to the work of these organizations compared to the resource dedicated to actual national safety work. Often important long-term issues can only be addressed through the international agencies but the political path towards the change may be long.

Trafi's mandates and responsibilities are naturally reflected in the legal framework. For example, reporting safety occurrences to Trafi is mandatory in several modes of transport.

1.2 Safety data, risk identification and risk assessment

A key department in the context of safety data and analysis was the (transport) analysis department which was a part of the "regulation and development" -division. This was the place receiving virtually all the operational safety data flowing into Trafi. This was also the home for analysis and risk assessment based on the available information. The analysis department had dedicated experts for the four modes of transport, working both autonomously and as a team, depending on the tasks. Other units and services would use analyses provided by this department in their own work. For example, auditors would ask for a specific analysis on a particular operator before going to perform an audit there. Analysis requests were also received from Trafi top management and from the ministry of transport. To help deal with the analysis requests, some standard analyses had been defined, specifying the depth of the analysis and the lead time. The analysis department would try to process the constant data flow in line with applicable regulations and agreed practices. For example, it would categorize the incoming aviation events using the ECCAIRS taxonomy. The department would also contribute to the yearly and quarterly reports by the agency. However, there was no holistic risk picture which would be kept up to date.

Each mode of transport had their own specific data sources, data types and challenges. The analysis work was also different from one mode to another. Appendix 5 contains the detailed results in terms of data sources and analysis activities for each mode of transport in 2013. The results in terms of data analysis and risk assessment can be summarized in the following way:

- Road transport: The day-to-day work of the road safety team in the analysis department consisted mainly of creating *status analyses based on indicators* (typically reflecting quantities of various types of accidents or fatalities) and producing *ad hoc analyses* as requested by Trafi management or the ministry.
- Railway transport: The analysis work concentrated on *maintaining safety indicators* both for Trafi and for EU level statistics. Additionally, a *yearly safety report* was made and many *ad hoc analyses*, which sometimes contained qualitative risk assessments typically on planned changes. The data sources did not provide a base for carrying out continuous quantitative risk assessment.
- Marine transport: The analysis activity focused on *maintaining the local safety indicators* and carrying out various *ad hoc studies* and contributing to the *yearly safety report*. There was no continuous quantitative risk assessment activity. Lloyd's List Intelligence does contain risk profiles and ParisMoU makes risk classifications on ships and ship-owners but this is far from a local risk assessment focusing on what would be the key maritime safety priorities for Finland. The 3-level safety indicator system had not been fully completed for the maritime domain, so the third level (causal factors) indicator definitions were still in progress.
- Aviation: Based on the various processes used, *events got classified* in various ways: severity class, occurrence category, event type, links to used indicators, responsible body, department or person. There was *no method for risk assessment* nor a model that would support such a method. As its key deliverable the occurrence analysis process led to results in the form of the *safety indicators*. The indicators count how many times different types of occurrences have taken place without trying to assess the risks involved. Different types of *analyses* were also performed and the topics for the internal review meetings were chosen and prepared. Periodical *safety statistics and publications* were produced and maintained.

It becomes clear that what was missing was:

- A systematic approach which would treat all the available information in a standardized way
- The capability to carry out risk assessment
- Consequently, the capability to present a holistic view in the safety risks within one mode of transport, and finally at the level of the whole transport system covering the four modes of transport.

Key decision makers presumably had their own holistic views of the risk situation, but such views were not built based on a systematic risk assessment process and they could not be made explicit and debated.

1.3 Decision making and taking action

Trafi's structure with the mentioned three organizational units formed a natural basis for decision-making within the functions but no easy solution for decision-making *across* the functions. Frequently, matters within one *mode* of transport overlapped several functional divisions. There was no dedicated decision-making body for such situations. In fact, the only level at which decisions could be made across functional divisions was the management team of Trafi. However, the management team meetings were obviously not meant for operational decisions. This introduced a handicap for decision making.

The knowledgeable experts for each mode of transport where spread among the three functions. To gather the expertise for analysis and decision-making within one mode, a Risk Review Board (RRB) existed in some modes of transport. The RRB participants found such work interesting but were frustrated due to the difficulty to transform RRB conclusions into decisions at Trafi level (and into action). The director general for the mode of transport sometimes implemented the RRB decision as his personal decision, but this was unusual, and the decision could still be questioned by the Trafi management team. The same cross-functional challenge would apply to following up decisions and on-going actions.

As there was not process nor methods to review risks in parallel in a holistic manner, the decision making took place case-by-case. All actions would need to be taken by the organizational units, within the three functions.

The way the agency took action was classic: regulatory work, authorizing or not requested activities and carrying out oversight of operators. Additionally, Trafi carried out continuous safety promotion through safety bulletins, reviews, seminars, etc.

2 Developments at Trafi during the TiTo project

2.1 Background and objectives

Safety management systems had been around for several years and the related idea of operators being responsible for their own safety was becoming the dominant paradigm. The role of the safety authority would be to ensure the good functioning of the operators' safety management system, instead of imposing compliance with prescriptive rules. In other words, the worldwide trend was to replace compliance-based regulation with performance-based measures. As a result, some authorities, especially in aviation, were already doing so-called performance-based oversight. The concept was that the authority assesses operators regularly and tries to get an idea on how safe their operation is and how well safety issues are addressed through the operator's own safety management system. The more the authority has confidence in the operator's own capability to manage its safety, the less oversight activity it would allocate on the operator – keeping more oversight resource available for the weaker operators. In practice, the assessment is usually carried out using a very simple point system which then splits the operators in different "confidence groups" (Booth 2012, Swedish Transport Agency, 2014).

Trafi's management wanted to go this way, but not only for oversight activities. The whole operation of Trafi would need to become guided by risks. The need for such a careful allocation of resource was

underpinned by the realization that Trafi would not get more resources. The operating environment was dynamic, introducing new challenges frequently and Trafi's resources were already somewhat stretched. It would need to use its limited resources in a way which would maximize its positive impact. This meant prioritizing issues, and the most important criterion for prioritization would be (safety) risk.

Consequently, Trafi wanted to obtain a risk management methodology which would help it allocate its resources intelligently in line with its mission across the four modes of transport and all organizational functions.

In 2013, the tendency to integrate the four modes of transport together as much as possible was very strong. This probably reflected the strong will to ensure that the previously separate agencies would not remain separated under Trafi's umbrella. Also for the TiTo project, the desire was to make one approach that could be implemented at the transport system level, so covering the four modes.

The overall objective of the project was to create the risk management framework for Trafi, which would enable the risk-guided modus operandi. The projected components of the project at the time were:

- Making sure all possible safety data could be captured and utilized
- Creating a holistic risk picture
- Identifying the safety factors which ensure the safety of the transport operation
- Prioritization based on risk, using the risk picture and the safety factors.

2.2 Methods

The researcher cooperated with Trafi very closely during the two years of the project. The approach was *participative*, i.e. the development work was done in constant interaction instead of just sending deliverables to Trafi from time to time. The researcher made 14 visits to Trafi, each lasting several days.

The initial analysis carried out in 2013 was summarized in Chapter 9.1.2. In parallel to that, both the development of the various components for the risk management framework and training sessions to Trafi personnel started. The researcher developed the components, trying to take into account the needs and constraints observed at Trafi. Initial versions were used at Trafi, experiences were discussed, and further refined versions were designed. Each mode of transport usually needed to be addressed separately.

Several trainings were delivered for Trafi personnel to support the new proposed working methods:

- Detailed training on risk
    - For members of the analysis department only
    - During first three months of the project, six one-hour sessions
    - Risk and ERC and SIRA according to the ARMS methodology
- "Safety and Accidents"
    - 3-hour training in the auditorium, 6 months after the start of the project
    - Cartesian-Newtonian view to causality and safety and the resulting methods including safety taxonomies.
    - Different models of accident causation
    - Evolution of risk assessment
    - What kind of safety data should be collected
    - Project content: how risk management should evolve
- "Risk-guided decision making"
    - 3-hour training in the auditorium, year and a half after the start of the project
    - What is risk-guided decision making, its goals, and how to get there
    - Holistic risk picture
    - Theory on black swan risks and on randomness
    - Resilience, complexity
    - How safety decision making could be organized in Trafi

- "Risk-guided decision making" (customized for the management team)
  - Decision making process and the use of the risk pictures
  - Balancing safety with other priorities
  - The black swan corner and resilience
- "Assessing Safety Cases"
  -
  - Special trainings separately for aviation and railways
  - How to assess the acceptability of a Safety Case proposed by an operator?

Regular steering group meetings provided access to the management team of Trafi and a good support from that level. The director of the analysis department acted as the primary contact person on Trafi side throughout the project, and the analysis department in general was the most exposed part of Trafi to the project.

2.3 Development and testing of NRMF components

During the first months of the project, the focus was on developing the required methods and capabilities at the analysis department to transform the incoming safety data into risk information. The training on risk (also covering ARMS ERC and SIRA) was a good basis for starting to create a new event risk assessment method for Trafi, customized separately for each mode of transport, but producing comparable results. The event risk assessment matrix for marine events was presented in Figure 18. The matrix for each mode was used as a printout, attached in front of the analyst when carrying our event risk assessments. The actual entry was made in a specific excel file.

Figure 26 illustrates such an excel file for railway transport events.

Event ID	Description	Indicator	Date	Activity	Org1	Org2	Equipment	Scenario	Severity	Human_pot	Envi_pot	Mat_pot	p_pot
OCC731		Opasteen vaiht	21-08-14	Matkustajaliikenne			normaali	matkustajan loukkaantuminen vo	d1	100	0	0	0.005
OCC732		Opasteen vaiht	26-08-14	Matkustajaliikenne			normaali	matkustajan loukkaantuminen vo	d1	100	0	0	0.005
OCC733		Ilkivalta	29-08-14	Tavaraliikenne			normaali	allejäänti	c2	10000	0	0	0.1
OCC734		Muu vaaratilan	16-08-14	Matkustajaliikenne			normaali	matkustajan loukkaantuminen	d1	100	0	0	0.005
OCC735		Muu vaaratilan	16-08-14	Matkustajaliikenne			normaali	matkustajan loukkaantuminen	d1	100	0	0	0.005
OCC736		Törmäys esteen	19-08-14	Matkustajaliikenne			normaali	suistuminen	e7	70000	0	0	1E-06
OCC737		Törmäys esteen	25-08-14	Matkustajaliikenne			normaali	suistuminen	e7	70000	0	0	1E-06
OCC738		Törmäys esteen	01-09-14	Matkustajaliikenne			normaali	suistuminen	e7	70000	0	0	1E-06
OCC739		Toimintä radan	18-08-14	Ratatyö			normaali	allejäänti	c2	10000	0	0	0.1
OCC740		SPAD	28-08-14	Matkustajaliikenne			normaali	törmäys	d2	25000	0	0	0.005
OCC741		Opasteen vaiht	28-07-14	Matkustajaliikenne			normaali	matkustajan loukkaantuminen vo	d1	100	0	0	0.005
OCC742		Kulkutien turva	02-09-14	Vaihtotyö			normaali	törmäys	d1	100	0	0	0.005
OCC743		Ilkivalta	16-08-14	Matkustajaliikenne			normaali	ikkunan rikkoutumisesta johtuva	e1	1000	0	0	1E-06
OCC745		Muu vaaratilan	16-08-14	Matkustajaliikenne			normaali	törmäys	d2	25000	0	0	0.005
OCC746			30-08-14	Matkustajaliikenne			normaali	suistuminen	e7	70000	0	0	1E-06
OCC747		Tasoristeyspuo	03-09-14	Muu			normaali	tasoristeysonnettomuus	b1	100	0	0	0.5
OCC750		Itsemurha	28-08-14	Matkustajaliikenne			normaali	allejäänti			0	0	0
OCC751		Muu vaaratilan	24-08-14	Vaihtotyö			normaali	törmäys	d2	25000	0	0	0.005

Figure 26. Event Risk Assessment entry matrix for railway events (extract).

The event description column contains a short narrative of the event (not visible in the figure to protect the data). The next column links the event to safety event indicators used in Trafi, so that this dimension can be used in statistics. The "activity" could be passenger traffic, shunting or track maintenance, which are each distinct operations from the safety management point of view, but could interact in real life. The next two columns allow tracking of organizations involved, so that risk data can also be viewed from this dimension. The "equipment" column allows recording what was the equipment in question

(train, locomotive only, track maintenance equipment, etc.). The "scenario" column records the scenario that the analyst considers the reference for the first question in event risk assessment: the most credible scenario for an escalation into an accident outcome. If the printed matrix is used, the chosen square is entered in the column "square" and the rest of the columns will be filled in automatically. If the analyst so chooses, the various severity and probability values can also be entered directly into the respective columns. There is a specific column for human, environmental and material losses/damage. The last column is for the estimated probability of the escalation.

More columns are needed to include the Safety Factors and the fact that some events may have actual outcomes, in addition to the potential outcomes. These are presented in Figure 27.

Event ID	+ Safety Factors						- Safety Factors						SEVERITY (of actual outcomes)				
Event ID	t1	t2	t3	t4	t5	t6	n1	n2	n3	n4	n5	n6	square_loss	Human_loss	Envi_loss	Mat_loss	p_loss
OCC731							25						a1	0	0	0	1
OCC732							25						a1	0	0	0	1
OCC733	14												a1	0	0	0	1
OCC734	71												a1	0	0	0	1
OCC735	71												a1	0	0	0	1
OCC736													a1	0	0	0	1
OCC737													a1	0	0	0	1
OCC738													a1	0	0	0	1
OCC739							24						a1	0	0	0	1
OCC740							28	26					a1	0	0	0	1
OCC741							25						a1	0	0	0	1
OCC742							24	1A					a1	0	0	0	1
OCC743													a1	0	0	0	1
OCC745							62	84					a1	0	0	0	1
OCC746	28												a1	0	0	0	1
OCC747													a1	0	0	0	1
OCC750													a2	10000	0	0	1
OCC751							24						a1	0	0	0	1

Figure 27. Columns for Safety Factors and already materialized actual outcomes.

The columns for positive and negative safety factors are used for noting which safety factors are relevant for the event. In the interest of space, short codes are used for the safety factors, e.g. 24 or 1A. The file is built in such a way that it is able to establish statistics associating the safety factors with the corresponding event risk values. If an event already had an actual outcome, i.e. a materialized risk, it can be noted in the last five columns on the right. Again, this can be done either by using the matrix or by entering specific values directly into the corresponding columns. For actual outcomes the probability value would always be one.

There were various experiments taking place in the different modes of transport around event risk assessment. The results of one such experiment are presented in Figure 28. In this case, events risk assessment was performed on a batch of 381 maritime safety reports related to the piloting activity. The key advantage of the accumulated results from event risk assessment compared to classical safety indicators is immediately obvious: looking only at the number of cases, technical failures seem to be the most important category; but the picture given by the cumulative risks shows that the maneuvering skills related events are far more critical. Already the cumulated risk of the technical failure events is much lower than the one of the maneuvering skills events. Moreover, a part of the risk related to the maneuvering skills comes from cases where the event risk was very high (red color).

The author developed the initial versions of the safety factors for each mode of transport based on the principles outlined in Chapter 7.2.3. Feedback was collected from experts in Trafi both individually and through specific workshops where it was possible to check that the set of safety factors was complete and that the used terminology was correct and precise. Use of the safety factors on real safety data by Trafi analysts also revealed opportunities for refining the safety factors. Throughout the evolutions it was made sure that the safety factors still reflected the original principles. As shown in Figure 27, it was possible to combine the allocation of safety factors to the event risk assessment and manage both in a single tool.

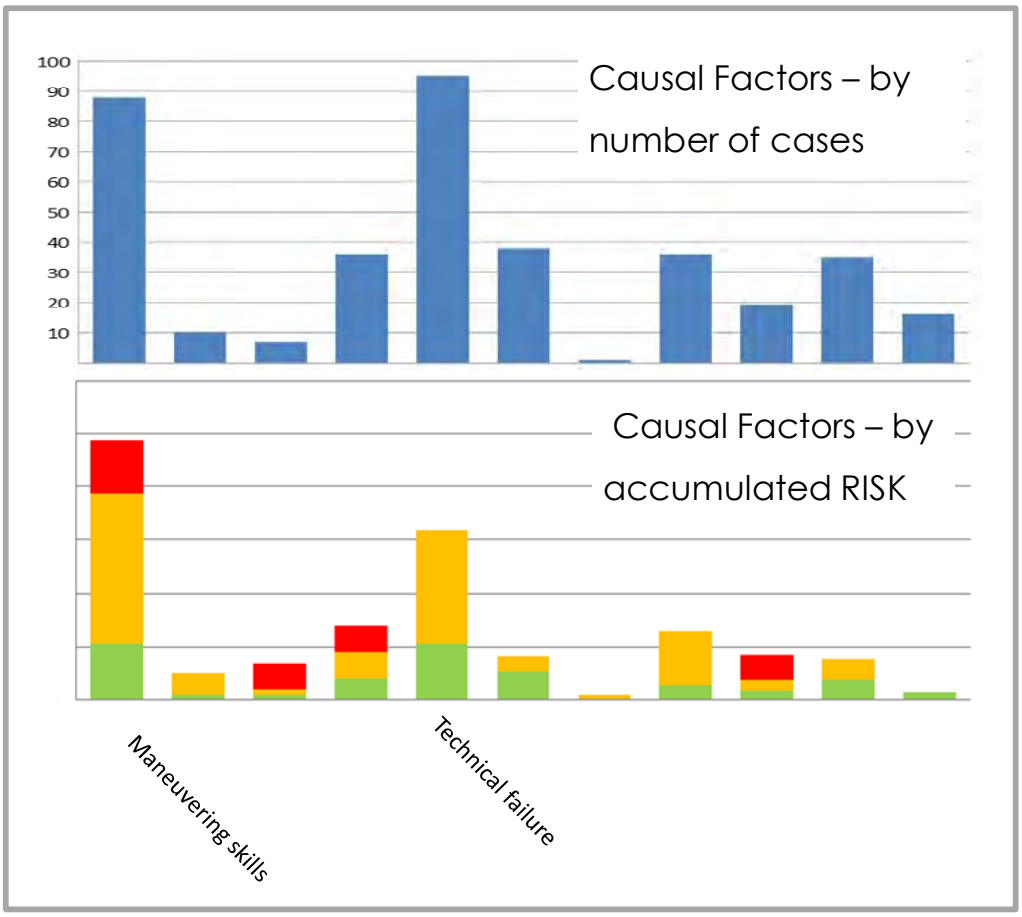


Figure 28. Example of early application of event risk on marine safety events. N=381.

During the development of event risk assessment, it was decided that the advantages of having a single scale for the different dimensions of severity outweighs the problems related to establishing the relationships and accepting the subjective nature of such a result. In other words, severities related to human fatalities and injuries, material losses and environmental damage would all be expressed with the same point system and on the same scale. There are some risk assessment matrices in the industry which feature multiple severity dimensions and could therefore potentially have offered guidance in bridging the different severity dimensions. However, it is easy to observe that while the same (often numerical) scale has been used for each dimension, running from min to max value, there has been no calibration between the different dimensions: e.g. whatever is the maximum consequence within a specific severity dimension finds itself at the same level with the maximum consequence of other severity dimensions and these are impliciply assumed to be of equal severity value. For example, in NPSA (2008, p.6), among others the following maximum severity consequences are presented:

- Incident leading to death (domain: impact on the safety of patients, staff or public.)

- Prosecution (domain: statutory duty/inspections)

This and other similar examples left the author convinced that the various consequences listed at the same severity level have not been carefully calibrated to really reflect the same level of importance for the organization in question. Therefore, rather than trying to use an existing reference it was decided to organize a specific exercise to get different types of outcomes (with different severity dimensions) assessed using the same scale, so that the "bridging coefficients" could be determined. This would also be more justifiable in making sure the results reflect the values within the time & place in question. As discussed in Chapter 7.2.1, the objective was closer to a *proof of concept* than trying to establish the "correct" coefficients.

The author designed a specific form on a single PowerPoint slide for carrying out the exercise (see Appendix 6). The center contains a vertical scale reflecting human lives. Different types of outcomes were placed on both sides of the scale, reflecting material damages, injuries and environmental damage. These outcomes were chosen to represent significantly different magnitudes of damage to cover the whole range of potential outcomes. The participants were asked to drag the pointer of each outcome onto the severity scale, so that the impact of the two losses would seem equal to the respondent (e.g. "how many human lives would you be ready to sacrifice to avoid this outcome?"). The exercise was carried out among the people working in the analysis department and the steering group of the project, totaling to 22 people from Trafi. Most people used the file and some a paper print. All participants were able to complete the form properly. The results showed that there was agreement among the respondents that material damages have little weight compared to fatalities. However, there was a lot of scatter in the results concerning severe environmental damage. Even if demonstrating the feasibility of this method was more important than the results, it is still interesting to observe that the highest-value material damage outcomes produce cost-per-life-saved values which are roughly an order of magnitude higher than the typical values - Vanem (2012) quotes several values below US 10M and an upper bound of US 15M. Allocating an existing budget (of the agency) between saving human lives and something else seems to be a very different decision-making situation than deciding or not to make a specific safety investment.

The so-called Polscale-table which categorizes different levels of oil damage, was used as a reference for the different severity classes for environmental damage. For injuries, there was a consensus in Trafi that severe injuries could be considered as severe as fatalities, as the human impact is at a comparable level, and the associated costs for the society are often even higher than for fatalities, due to expensive long-term treatments. Obviously, a much larger sample of respondents will be needed to build a representative answer even at the level of Trafi, not to mention building a national consensus.

The ARMS SIRA was used in Trafi during the project, and a specific in-house procedure was created for the method. Gradually it became clear that safety issues should be placed in the risk picture along with other risk elements.

As the general perception at Trafi was that the actions taken by the agency tended to follow the old compliance-based paths, an effort was made to brainstorm as many action types as possible. Inputs were collected both from people in Trafi through interviews and workshops, and from the author. The aim was to collect a list of *possible* action types, including both classic well-known action types and unconventional ones (*without* judging which ones might be *recommendable*). The brainstorming produced the following items:

CLASSIC MEANS

- Regulation and recommendations
    - The trend is to reduce regulation and give more responsibility to the operators
    - Law changes require preparatory work
    - Trying to influence rulemaking at EU and global levels.
    - Trafi has traditionally refrained from giving *recommendations*
- Oversight

  - Audits and reacting to findings
  - Inspections
- Authorizations
  - Maintaining the criteria for authorizations and refusing when the criteria are not met.
  - Guiding the operator to carry out more detailed analyses
  - Limiting the authorizations/permits of an operator
  - Cancelling an authorization/permit. This would be a major intervention, as it would typically cause an immediate stop of the operation.

OTHER MEANS

- Direct influence on the operator
  - Meeting the top management regularly
  - Meetings between the safety department of the operator and a relevant team in Trafi (e.g. the analysis function).
  - Written or verbal contact to the operator. Just to acknowledge that Trafi has taken notice of an issue may have an influence.
  - Request for the operator to explain how the operator is handling a specific issue.
  - Sanctions: fine, warning, notification, threat of extra requirements/limitations, threat of interrupting/stopping the operation.
  - Making all the above actions public (only if legal).
  - Stressing the critical issues during audits.
- Indirect influence on the operator
  - Convincing another transport-related agency to take action or taking mutual action.
  - Cooperating across the political system, also with agencies outside the transport scope.
  - Influencing important stakeholders: e.g. driving schools, vehicle inspection stations, insurance companies, etc.
  - Meeting other stakeholders. The meetings can be informal.
  - Facilitating forums for exchanging good practices, e.g. between operators.
  - Providing good tools for users/customers and for the safety work at the operators.
- Showing direction, guiding, motivating
  - Establishing national strategies and visions.
  - Producing and distributing so compelling information that others have no choice but to start acting accordingly. This way, Trafi gets a directing role.
  - Trafi could become the coordinator in some topics, gaining positive influence.
  - Initiating a public discussion on the topic in question
  - Involvement with public interest groups, thematically.
- Educating and informing
  - Conferences and other events, specific campaigns, through stakeholders like flight instructors, during an audit.
  - Online courses for professionals.
- Data and information
  - Producing reliable and useful data for others, e.g. for the needs of design and development.
  - Better collection of safety data: creating methods and tools and promoting them.
  - Research. Bringing safety science to the operators.
  - Developing voluntary online safety reporting.
  - Promoting the safest roads, e.g. through GPS.
  - Crowdsourcing for collecting new kinds of safety information
- Other
  - Encouraging better practices
  - Specific focus on the weakest operators, trying to make them evolve.
  - Highlighting identified problems to the professionals who are closest to the them. They may be best positioned to find the solutions.
  - Asking the professionals/operators how Trafi could best help with a given issue.

- Channeling funding for safety projects.
- More radical and novel laws and regulations, e.g. legislation on young drivers in California. (see e.g. http://teendriving.aaa.com/CA/supervised-driving/licensing-and-state-laws/)

The objective in making this list was not yet to identify interventions specifically suited for complex systems but rather to remind people of the possibility to use novel means. Nevertheless, the list does contain some means which resonate with complex leadership, like facilitating forums where operators can share experiences, producing compelling information and highlighting problems to professionals and letting them find the best solutions. In any case, the list can give ideas for interventions.

Once the risk picture and the risk workshops had been developed as a part of the new process, both were tested with some example topics. Experts from different parts of the agency where gathered together to build knowledge on a chosen topic and the resulting discussions and identified scenarios were recorded. The later stages of formulating intervention proposals and challenging them were also tested.

Figure 29. Trying out the Ritual Dissent method. A presenter from another team gets feedback, facing away from the team.

The so-called ritual dissent method was used. This method developed by David Snowden's company Cognitive Edge builds on diversity and active listening. Several teams develop intervention proposals. A member from each team rotates to the following team and has a few minutes to explain the proposal to this neighboring team. The receiving team is not allowed to talk in order to benefit from active listening. After this, the presenting member turns around and receives critical feedback from the team and this time the presenter must stay silent. Facing away from the team should help make the feedback less personal. After a few minutes of this another rotation can take place and after a few rotations the

presenters return back to their original teams to share all the feedback received. It is possible to configure some of the teams to defend specific priorities like environmental sustainability. If all proposals need to rotate through all the tables representing different priorities than the final proposal should have a good chance of striking and acceptable balance between the different priorities.

The plan was to run more risk workshops and test-run the decision-making stages before the end of the project, but that did not materialize due to the re-organization of Trafi which had a major impact during the last months of the project. Trafi was left with a specification of the new risk management framework and most of the components either already running at least in some modes of transport or having been tested. In terms of safety data inflow, aviation had sufficient data but needed time to process a large backlog before adding any new processes, and the first useful safety data for railway transport started gradually flowing to Trafi from the biggest operator towards the end of the project. Processing the railway data with the above excel tool was immediately adopted, introducing both event risk assessment and safety factors. At the time or writing, the database contains over 1300 events. The data situation for maritime and road transport was different: there was no regular data flow for the former, and the latter only got data on losses – and the injury data was not yet available.

A separate project had been created in Trafi for starting organizational profiling of operators. This naturally had a close link to the organizational safety factors discussed in Part II. This link was recognized by the management, and the idea was to gradually merge the results from organizational profiling to the overall risk management process.

Figure 30. The author on the bridge of a ship during the resilience capture exercise.

The capability of the safety factors to help recognize operational resilience was explored at the end of the project by spending a day on the bridge of a large cruise ship. The maritime safety factors were used as a basis for observations and discussions with the crew on the bridge. The experience was very encouraging. It was possible to identify many resilience-enhancing practices in a short time. For

example, manual controlling modes were used in order to maintain the related skills, and due to narrow passages the times and locations where two ships would meets were coordinated between the different operators, especially when a ship got delayed from its normal schedule. The conclusion was that the safety factors provide a promising functionality also in exploring the Work-As-Done and especially in identifying the positive elements which are typically hard to detect, as discussed earlier.

3 Implementation of the developed process at Trafi

In view to the adopted paradigm of a complex adaptive system, care had to be taken in terms of how the results at Trafi would be assessed and described. This had at least two practical consequences. First, attention should be paid to all kinds of evolutions in the system during this time and not only the implementation of the new framework. Such evolutions would interact with the development of the framework and vice versa. For example, an aviation accident with eight fatalities occurred during the project and resulted in the director of the analysis department being nominated to lead a special project on leisure aviation safety, as her full-time activity. This impacted the project both through the analysis function and through the fact that this person was the key contact person for the project on the Trafi side. Secondly, the complexity of the activity and the idea that knowledge is spread all over the system meant that typical questionnaires would be too superficial to capture many of the interesting underlying rationales and insights that people in Trafi would be able to offer if given the chance. Therefore, it was decided that the results would be gathered through interviews even if their analysis would be more challenging.

In the interest of objectivity, a person from Trafi was involved in carrying out the whole interview exercise together with the author. The interview questions proposed by the author were commonly reviewed and approved. The list of people to interview was established together and the interviews split between the two. All interviews were recorded and stored. After all the interviews had been done, both interviewers listened to all the interviews and wrote independent summaries/conclusions question by question. These summaries were then compared, discussed and merged into a final agreed version.

The five questions in the interviews aimed at getting insights on:

- What has been implemented in Trafi in relation to the new risk management process?
- Have the mindsets or paradigms at Trafi evolved thanks to the exposure to the project, the related trainings and new methods?
- What else has changed in Trafi or its operating environment between the project start (April 2013) and now (early 2017)?

The following people were interviewed:

- Heli Koivu, Chief Adviser to DGCA (former Department Director – Transport Analysis)
- Ilkka Kaakinen, Chief Adviser - Aviation (former Head of Unit, Safety Analysis)
- Ville Autero, Head of Unit, Safety Analysis
- Kirsi Pajunen, Chief Adviser to Director Rail transport (former Risk management coordinator- Rail transport)
- Ville Vainiomäki, Special Adviser, Safety Analysis, Rail transport
- Tapani Maukonen, Special Adviser, Safety Analysis (Risk management coordinator- Aviation)
- Valtteri Laine, Special Adviser, Safety Analysis, Maritime (recently left Trafi for Helcom)
- Riikka Rajamäki, Special Adviser, Safety Analysis, Road transport
- Kristiina Roivainen, Development Manager, Organization Services

The summary of the results, created with Mr. Ilkka Kaakinen 11th April 2017 is presented in the next section. It is followed by an independent assessment of the risk assessment method in Trafi for aviation and an overall analysis of the Case Study.

3.1 Results from the interviews

1. **Since the project, what significant has changed either in Trafi or in its operational environment?**

- The **strategic scope** of Trafi has enlarged and there is more focus than before on priorities such as commercial transport activities and environmental sustainability. Today safety is one of many priorities and focus areas of Trafi.
- The strong desire in 2013 to standardize across all transport modes and to focus on the whole transport system as one, has transformed into an approach where **each transport mode has more freedom** to embrace its specific features.
- Trafi has gone through several **reorganizations**, including a major one, where the current matrix organization was created, and all managerial positions were redistributed through an application and selection process.
- The updated Finnish Aviation Safety Programme (**FASP**) was established and approved by Trafi in 2017, containing a Finnish aviation risk management process inspired by the Tito-project.
- Helsinki **metro** traffic has become part of Trafi's scope, and the **tramway** operations will soon too.
- Many smaller distinct **improvements** have emerged: e.g. cooperation with other agencies has accelerated, the safety reporting system of the Finnish Transport Agency (LiVi) is running now, the main train operator has become more active in safety work than before, tracking severe injuries in road safety has become possible, Trafi has been given the lead role in Finnish road safety coordination, and managing the flow of safety reports in aviation has become possible thanks to digital reporting.

2. **Has the project changed ways of thinking/working?**

- The absorption has been very different for **different transport modes and people**. Biggest positive change was in aviation and among people with a good base knowledge and operational touch. In several areas, there has been no visible change.
- Majority of Trafi activity is **still compliance-based**, and the related mindset is still very much present.
- The Trafi experience suggests that change in working methods **requires strong support and push from the top management**, which was the case in aviation.

3. **Have the working methods changed concretely?**

- **In aviation, the whole safety process** (FASP) changed profoundly and integrated the risk-informed paradigm. The process includes e.g. the risk picture, risk workshops, decision making panels, and a supporting software has been specified.
- Incoming **Railway safety reports** are entered in a specific excel database and they are subject to event risk assessment and identification of the related Safety Factors, in line with the Tito-specified working methods.
- Elsewhere in Trafi, Tito-methods have **not been implemented systematically**. Some of the reasons offered include: low resources and compliance-based tasks being prioritized higher, no systematic incoming data flow (maritime), ERA-mindset focusing on operator self-management through SMS (no risk assessment at agency), specific road safety traditions and traffic being generated mainly by private people rather than organizations (operators).

4. **What is the experience of the new working methods?**

- For the implemented processes and methods, the experience is positive and **encouraging**.
- Methods have been **customized** to Trafi's needs (e.g. the organization), and early **feedback** has been taken into account rapidly.
- There are **concerns about the workload related to Event Risk Assessment**, which has delayed its implementation in aviation.
- **Not everybody** in Trafi believes in the value of the new working methods.

5. **Does the developed Tito process help Trafi in fulfilling its mission?**

- **Yes**, when implemented.
- Trafi's mission has evolved during the last years. It is considered that the Tito-process is in line with both the older safety-related mission, and the more recent mission which stresses a data-driven modus operandi.

3.2 Independent study of the Trafi risk assessment method in aviation by the Technical Research Centre of Finland

In February 2017, the Technical Research Centre of Finland (VTT) performed an independent assessment of the risk management method used for aviation in Trafi. The aim of this study, commissioned by Trafi, was to support the on-going implementation and refinement of the methodology. VTT studied the "as is" method the way it was implemented at the time, knowing that the further development was in progress, especially in terms of creating a dedicated software product to support the process.

The scope of the process which was assessed by VTT covered:

- Practically only the aviation domain as the implementation was much more complete than for other modes.
- Event risk assessment, the related matrices and excel tools
- The risk workshops
- The excel-prototype for presenting the risk pictures
- The way to transform different severity dimensions into comparable points
- Safety Issue Risk Assessment (SIRA) which was at the time still performed in line with the original ARMS method.

It can be said that what was assessed by VTT was a prototype in terms of the methods, tools and organizational processes. Neither the way to present risks in the risk picture, nor the way to run the workshops were identical to the NRMF as described in this dissertation. However, the key features were already present at least as prototypes.

The conclusions of the study were the following (VTT 2017, pp.18-19)):

- The method covers properly all the stages of risk assessment
- The method has been well customized to the specific area of application
- Potential severities have been studied from relevant perspectives
- Probability classification takes into account the fact that most threats under study are rare.
- Calculations for the total severity and probability in the used excel tool are correct.

The identified areas for improvement include:

- There should be detailed guidance on how the descriptions of the threats and scenarios under study are done.

- Implementing techniques which counteract typical human biases and heuristics.
- More structured method for making the initial risk assessment of a threat, so that the rationales get recorded.
- Review of used values for severity points for the different severity dimensions, including more precise descriptions of material damages, environmental damage and smoothness of traffic flow.
- More support for assessing the probability of a scenario
- Presenting the uncertainty of estimates using symbology.
- Cost-benefit considerations could be developed through the estimation of residual risk.

4 Analysis of the Trafi Case Study

Perhaps the first thing to say is that Trafi and not least its top management demonstrated an impressive level of courage in engaging in a project which was bound to propose disruptive changes of paradigms and key processes. This was far from what people would typically expect from a state agency.

As a summary of what had been implemented by Trafi during the initial two-year period and later before the writing of this dissertation, one can refer back to the 12 process stages listed in Chapter 7.6 and to the flowchart of Figure 25:

- During the initial development period of two years, the event risk assessment and safety factors capability was put in place in the analysis function for all the transport modes. The way of working in risk workshops was also introduced and tested.
- The risk workshops were fully implemented within the aviation domain. This means that the process stages 1-4 were covered. The analysis function even managed the development of a software for building the risk picture prototype in an electronic format.
- Concerning the stages 5-6, organizational profiling was launched during the initial two years as a separate process and since that time the practice has been maintained and it has evolved at Trafi. The author is not aware whether cross-analyses have been carried out.
- Concerning the risk treatment part (stages 7-10), it is known that the risk workshops in aviation are operational, decisions are taken, and inventions are implemented. It can be assumed that proper follow-up is being carried out. However, these are stages which have been put in place gradually in the last years (after the 2-year development period), within the aviation domain, and the exact nature of interventions used so far is not known to the author.
- The preliminary analysis steps (stage 11) are being assured by the analysis function. The capability was created for all transport modes. Obviously, the most active domain is aviation thanks to the actively running risk workshops. The railway domain has also produced a systematic event risk assessment & safety factor analysis of all incoming event reports.
- Due to the asymmetric implementation of the process across different modes, a common risk picture (stage 12) has not been created so far.
- Referring to the flowchart, all the functional units (e.g. analysis function) are in place for ensuring the flow. The aviation process goes through the whole flowchart, with several specific risk workshop teams and dedicated electronic risk pictures. It is not known how often external people are involved ("communication and consultation"), and how elaborate the feedback collection methods are ("monitoring and review"). The other modes have a partial implementation and the situation can be expected to evolve. Concerning communication, the whole process has been made public by Trafi both as a part of the initial development project and within the Aviation Safety Programme (see below).

As the results of the interviews indicate, the implementation and the impact of the new paradigms have been quite different from one transport mode to another. Several factors have been proposed in the interviews that could explain this situation and it is impossible to know how important each factor is. For example, despite the lack of traditions for this type of risk analysis and despite the lack of pressure at the international level, would simply adding more resource facilitate the quick adoption of the new approaches in the railway or maritime domains? The experience from aviation suggests that even with

an international (ICAO) requirement, European requirement (EASA) and long traditions of data-driven safety work, top management support and hard work is required to get such a new process implemented at the agency. The situation has paradoxal aspects: there are no resources for implementing the new risk management process due to "mandatory" tasks (e.g. activities related to IMO and ERA), but on the other hand the process would help focus the scarce resources on the most important tasks.

Concerning aviation, Trafi published its new national Aviation Safety Programme 13th March 2017 (Trafi 2017). This plan has key parts of the proposed risk management framework embedded in the official national process, including the risk workshops, the risk picture and the so-called risk panel which is the decision-making body. Figure 31 illustrates the way scenarios are currently visualized in the risk picture. It is clear that the presentation format is in line with the proposed NRMF, but in lack of an adapted software, a temporary compromise is in use. For example, the scenarios are currently not presented as rectangles but as points. In any case, the Safety Programme gives the official regulatory framework under which the detailed process can be run fully in line with the proposed NRMF if this is considered desirable. The aviation domain has been divided in sub-topics and each sub-topic has its own risk workshop. The process has started running with the risk workshops having monthly meetings and the initial feedback from participants is positive. Participants are aware that the first year is particularly demanding because this is the period when the first versions of the risk pictures are gradually built up.

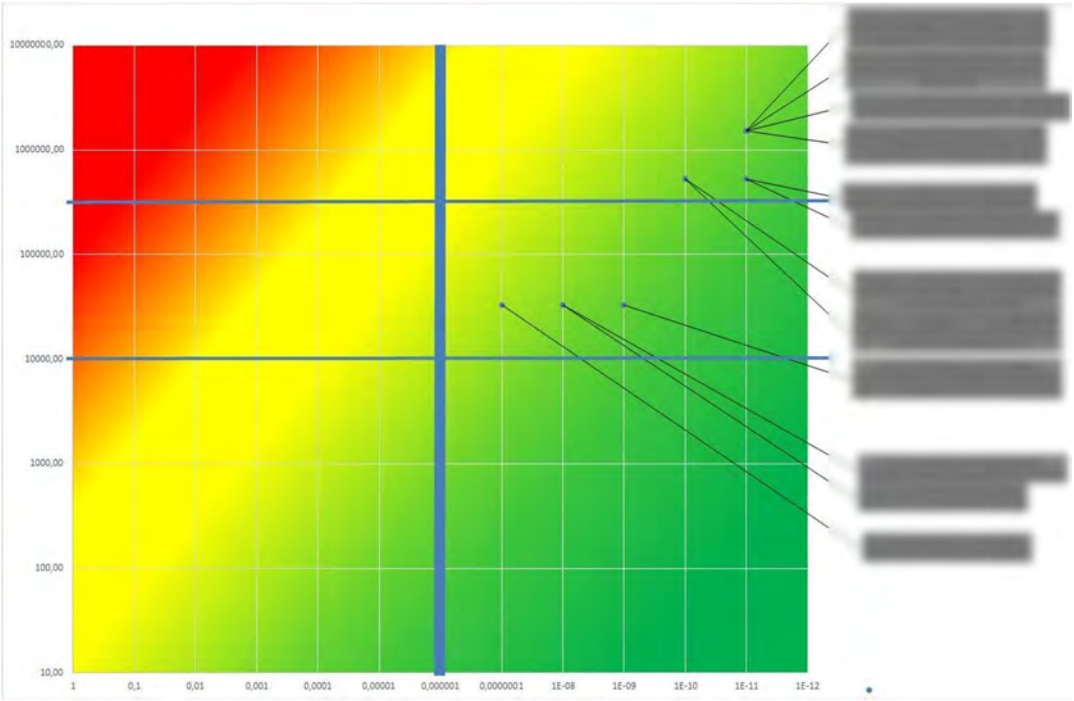


Figure 31. Example of the current temporary solution at Trafi to visualize aviation scenarios in the risk picture.

The results of the interviews also provide a delicious example of a complex adaptive system in action. The answers to the question 1 show some of the changes and evolutions that have taken place since 2013. Most strikingly, the very principle underpinning the project at Trafi – i.e. treating the four modes of transport through a single integrated risk management process – had become obsolete. The reasons for this change can only be guessed. Had the integration of the formerly separate agencies been so successful that the management was more relaxed to let the different transport modes live a more independent life? Were there concrete experiences which suggested that integration might not be the ideal solution? Or did some specific transport modes resist the proposed safety management paradigms? Whatever the reasons, the logical consequence was probably that there was no strong top management push for establishing the integrated process and the integrated risk picture rapidly.

The reorganizations were also typical examples of *organizational adaptations*. The second re-organization implied that a new organizational structure was set up and then all current employees had to re-apply for the (new) positions. This was a massive process during the last quarter of the project and impacted everything in Trafi at that time. The *adaptation at the individual level* meant that several people left Trafi, in addition to the ones who changed positions internally. There have been around 10 people in key positions in relation to the RMF who have either changed positions or left Trafi since early 2013. The analysis department has become smaller, and for example railway analyses are ensured by a single analyst. Adding the current expanding strategic role of Trafi – bringing in analyses of impacts of planned regulatory changes etc. – it is very difficult to add new tasks or processes. Even in the aviation domain, where there are more resources, there are doubts about the capabilities to run all incoming events through the event risk assessment, as indicated in the interview results.

The positive message is that the adopted parts of the new process seem to deliver the expected benefits. The ad hoc style of treating issues one by one is being replaced by a systematic process where all threats are reviewed systematically and holistically and studied with the benefit of all the internal expertise. The feeling in Trafi was that the new approaches help Trafi fulfill its mission – when they are implemented. Besides aviation, another concrete implementation is the event risk assessment / safety factor analysis in the railway domain, introduced in Chapter 9.2.3. The process is on-going and the standard statistics can now be presented with the new currency – riss – in addition to the more primitive event *count* which is the currency in classic safety indicators. Figure 32 gives an example of one of the charts which are produced automatically by the excel application used. Once again, it is obvious that the risk dimension puts the indicators in a completely different priority order than the simple event count. In the railway domain there is currently a mismatch in Trafi between the analysis department, which is producing such advanced risk-guided material, and the rest of the organization which is still following the classic state-of-the-art process without holistic risk pictures. A key difference with aviation is that the international roof organizations in aviation require a State Safety Programme with a risk management process, whereas the approach of the European Railway Agency (ERA, now replaced by the EU Agency for Railways) has been leaning on self-assessment of the operators, with no explicit incentive for the agencies to run risk assessment processes. This may change in the future with the new set-up of the agency.

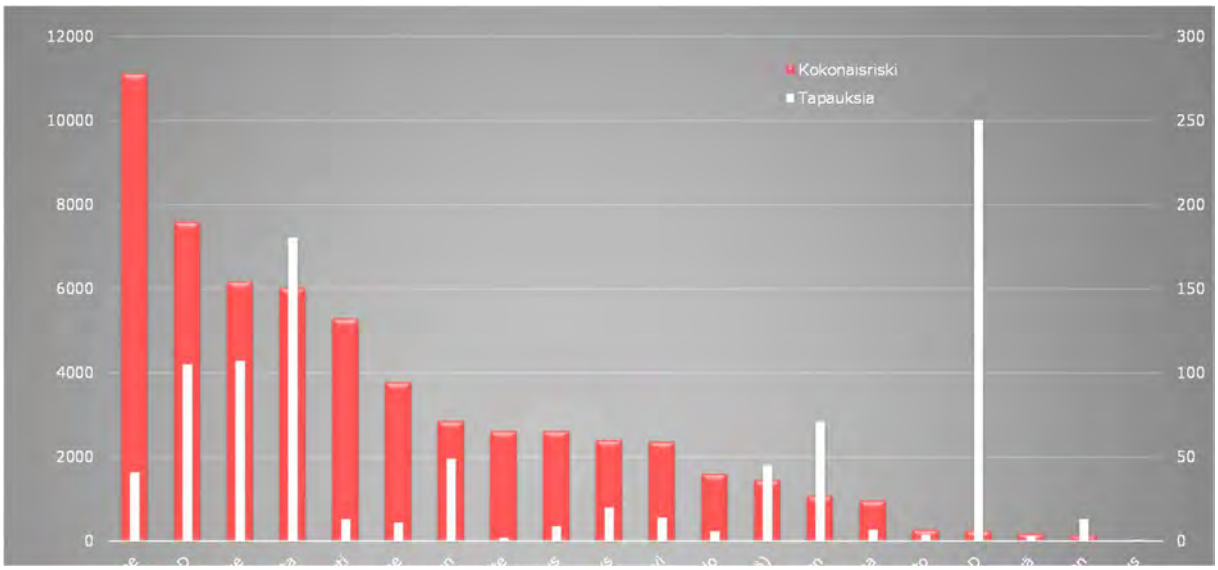


Figure 32. Example of a chart with standard railway safety indicators, with the results given both as a count of events (thin grey bars) and cumulative event risk (thick red bars).

The relationship with risk seems to be quite different in road transport where the work focuses virtually entirely on materialized losses – above all on fatalities. Accident statistics and studies provide the safety data and there is no perceived need (nor the resource) for a similar risk-guided approach as adopted in aviation. It is true that for road traffic alone, the interests of such an approach could be limited. However, getting them into a common risk picture with the other modes of transport could be highly interesting. The hard data on fatalities would be very visible and the road traffic ecosystem could be the most promising for experimental/adaptive interventions. In the maritime domain, Trafi does not yet have a data flow and the modus operandi has not changed much.

The summary of the interview results was sent to the Director General of aviation in Trafi, Mr. Pekka Henttu, who is naturally also a member of the Trafi top management team. His comments were the following:

- It is true that today a narrow-minded focus on safety alone is not enough for safety enhancement. The agency needs to maintain a holistic view, including aspects such as safety, security, sustainability, commercial aspects, punctuality, reliability, speed, etc. These can be considered the performance factors of the transport system. They are often in conflict with each other, and the agency needs to address these conflicts in a balanced way.
- The various organizational changes in Trafi have aimed at growing and developing the agency. Strategic development is protected by keeping it as a distinct activity.
- The Transport Agency (LiVi, responsible mainly for the infrastructure) will be re-organized and Trafi will recover some of its current tasks.
- The TiTo project had a major impact. The top management has understood the development needs clearly. There will be focus on the improvement of processes and tools (already started at TiTo). Mindsets are important. TiTo supported positive change there too.
- The role of top management is very important in changing ways of thinking and working methods.
- The proposed TiTo process definitely helps Trafi in fulfilling its mission!

The content of the VTT-study was supportive. As can be seen, all the points presented as areas for improvement have been addressed in the NRMF of this dissertation.

Going back to the shortcomings of current risk management methods in the industry, listed in six points in the end of Chapter 6.3, the Trafi experience shows, that the NRMF can overcome these problems, not only in a scientifically robust manner, but also in a practicable manner:

- The risk workshops introduce a natural and credible opportunity for *knowledge-building* on all risks which need to be assessed.
- The *knowledge-dimension* is emphasized and its assessment is required as it needs to be highlighted in the risk picture. *Low-probability risks* are treated irrespective of their probability so they are not ignored.
- The Agency can use the *existing risks in the risk picture as benchmarks* for acceptable risk levels and/or *reason in terms of alternatives*, rather than try to set hard numerical limits for acceptability. This gives a practical solution to the unspoken impossible job of specifying acceptability limits for risks.
- The point-system, also used at Trafi, enables giving more points to high-priority items and thus *enables implementing priority lists*, such as the one of EASA.
- The existence of the risk picture gives *consistency to decision-making compared to a case-by-case approach*.
- Building up *alternative interventions and refining them with a carefully planned internal critique* process has been trained and trialled at Trafi. Interventions adapted to a complex system have been trained.

All the experiences with Trafi give confidence that the developed methodology is implementable without major problems. Certainly, at the time of writing the dissertation, different parts of the process had been implemented at different levels of depth, and an organization may always choose to implement

or not to implement various components, the way it wants. However, unlike before, the method itself will no longer be the barrier to proper safety risk management. The NRMF is implementable and it overcomes the listed shortcomings. Virtually all components of the developed NRMF have been at least tested at Trafi. The early stages of event risk assessment and safety factor analysis are running continuously in the railway domain, and the later stages are being implemented under the aviation safety programme as a standard continuous process.

The implementation at Trafi is clearly only partial and in progress. It is impossible to know how the situation will evolve in the future, and that will be defined by many factors, only some of which are related to the NRMF itself. Review of the past and imminent changes affecting Trafi – including the new work split with the Transport Agency – underlines the statement that there are "no controlled experiments in complex adaptive systems". The overall feedback on the implemented components is encouraging, but obviously, if the aim was to collect quantified data about the implementation of the NRMF, it would be a separate, resource-heavy exercise to carry out, and preferably earliest some time in 2019 so that Trafi has had enough time to implement the different parts of the process.

How vulnerable is a risk management process like the one developed here to lacking and changing resources? Due to the required knowledge level in terms of the aviation system and its risks and to the accumulation of this knowledge over time as part of the process, losing a key person within the process can be a significant loss. The same applies to the knowledge of the process, methods and tools used and the capability to lead the people through the different stages. Are there ways to make the process more robust against resource problems? The fact that the repetitive stages of the process must be low on resource has already been mentioned, and this applies above all to the event risk assessment. A second important point is the support provided by adequate software tools. These can lower the resource requirements, help structure the process and make it more understandable to people and speed up the various stages. In addition, robust tools can help cement the process in a more solid manner into the DNA of the organization. Finally, an even higher level of protection is provided by making the process the official regulatory process that the organization must follow. This is exactly what was done in the aviation domain at Trafi.

Chapter 8 demonstrated that the developed NRMF complied with the requirements created based on the literature review. Its added value compared to existing frameworks was also demonstrated through a comparison. Adding to these the real-life experiences from the Finnish transport safety agency gives confidence that the proposed work is implementable and can bring true added value in the activity of transport risk management.

Conclusions

The starting point for this dissertation was the observation that current risk management frameworks have not taken full benefit of the latest research. The methods in use today tend to present risk simply as the combination of probability and severity, without explicit focus on managing the related uncertainties and the kind of surprises that could escape the state-of-the-art scrutiny. Similarly, on the safety management domain, the Safety-I paradigm of "safety as lack of negatives" still dominates the emerging Safety-II paradigm where safety is seen as "presence of positives", e.g. of resilience. In treating risks, there is hardly any evidence that the level of complexity of the system would be recognized and that the intervention strategies would be adapted accordingly. It would be difficult to reach an ideal level of risk identification, assessment and treatment with such handicaps.

The research question was defined in relation to taking benefit of the latest scientific understanding of the key concepts: what kind of risk management framework should be used for managing transport risks if the modern risk perspectives and the latest understanding of safety were embraced, and if the transport system was considered a complex adaptive system?

The area of application is transport and more specifically the role of a national transport safety agency. This introduces many further challenges. One has to deal with four modes of transport each coming with their own traditions, methods and safety data types. Accident severities, probabilities and the associated uncertainties range from very small values to very high values. The transport system is very complex and embedded both in the national political ecosystem and in the international framework of companies, organizations and regulatory bodies. Safety must be considered together with other priorities such as security and cost. Risk treatment must be effective in this highly complex system, yet the overall process must be light enough so that it can be run continuously with acceptable use of resource. Finally, the paradigm changes associated with a new risk management framework addressing all these challenges could introduce a significant psychological barrier for the acceptance of the new framework.

Through the literature review, specific requirements for the new risk management framework were specified. The requirements address all stages of risk management but also the context of application. No evidence could be found of any existing risk management framework which would meet all the requirements. This confirmed the validity of the research question.

The new risk management framework was developed based on these requirements. A rich cooperation with the Finnish Transport Safety Agency was established and the agency has implemented, and is in the process of implementing components of the new framework.

The results of this research show that it is indeed possible to create a risk management framework which meets the given requirements. The proposed framework is able to use existing safety data and virtually any type of information as inputs and specifies a process of knowledge-building around the identified risks. Central to the framework is the so-called risk picture which present a holistic view of the various threats, scenarios and experienced events and supports risk evaluation. The Strength of Knowledge behind the various risk elements is an important consideration and is made visible in the risk picture. Surprises (e.g. black swan risks) also get their own specific treatment. Risk workshops are used to structure the work in risk analysis, evaluation and drafting the interventions. Due to the complexity of the system, adaptive intervention strategies such as experiments are embraced. While people, rather than information technology, are counted on for gaining increasing understanding of the system and making difficult decisions, countermeasures are put in place to limit the influence of potentially negative human biases, heuristics and behaviors.

Direct validation of the proposed risk management framework, based on improving safety performance, is not possible due to the (fortunate) scarcity of accidents in such safe systems and the intractability of

causes and effects in a complex system. Consequently, the developed framework was validated in three ways. First, the compliance with the developed requirements was reviewed by explaining how each requirement has been addressed in the framework. Secondly, the compliance of other current risk management frameworks with the requirements was reviewed in a detailed table. It is proposed that current frameworks do not address the complete set of requirements. The industry methods are out of date in terms of the risk perspectives. The more recent methods which embrace the modern risk perspectives generally do not exhibit the capability to use a flow of operational safety data as input. A common limitation for all current methods is that they don't specifically question what kind of system one is dealing with (e.g. complicated/complex) and propose suitable intervention approaches. The conclusion is thus that the developed framework is going beyond the current RMF's.

The third validation component was the case study on the Finnish Transport Safety Agency, Trafi. This case study aimed at illustrating to what extent the various components of the framework are implementable and produce useful outputs. The case study involved a real, living organization in a complex system. This set-up is therefore far from laboratory conditions, at the mercy of the constant change and adaptation taking place in a complex adaptive system. Indeed, in the four years, the agency had changes in its strategic objectives, in its scope, its input data, its organization (several times) and its people, to name just some examples. Witnessing such change right in the center of the system for which the framework was designed, underlines the importance of having embraced the complex adaptive system concept. Furthermore, the fact that two years after the end of the active cooperation the implementation has been very different in different modes of transport shows how much the prevailing paradigms and traditions influence the adoption of such new approaches.

Due to the organically changing, complex nature of the system, the implemented method will never be exactly identical to the NRMF as it is theoretically described: the implementation at Trafi was partial and *work in progress*. Therefore, in addition to accepting the impossibility of direct measurement of the impact of the new method, one also needs to accept the fact that what was implemented and observed was not identical to the method assessed in the table of Appendix 4. Despite this, the case study offers a valuable component of validation. The experience with Trafi makes a credible case for stating that key elements of the NRMF are implementable: the risk workshops, the risk picture and the overall process. The demonstrated implementation success and positive results through the interviews suggest that the NRMF can be expected to be implementable and perform as designed. Trafi was able to put in place the necessary processes and organizational structures, and people were able to adopt their roles and embrace the work in risk workshops. Despite not having a fully developed software tool to support the process, enough software development was carried out to at least be able to present a simple risk picture electronically. Trafi had been able to implement the method to the extent it had desired. It also received an external endorsement of the method through the independent study of the risk management process within aviation, carried out in March 2017 by VTT.

Based on the multiple components of validation, the conclusion is that the developed risk management framework brings significant benefits compared to existing ones, proposes a full process from the beginning of the cycle (data capture and risk identification) to the end (adaptive interventions) and is implementable and adapted to the real-life requirements like data capture and resource and time efficiency. The conceptual basis for the framework is in the body of existing scientific literature. New contributions to this body were also made, above all the new concept of *riss*, which has application in risk identification and analysis.

What can one do with the new proposed framework that was impossible or very difficult before? Most of all, one can see the integrated risk picture showing all risks side by side, so that priorities can be applied coherently rather than tackling risks separately, one after another. It is possible to study alternative interventions with their advantages and disadvantages, taking into account the risks, uncertainties and the multidimensional costs of the interventions. All this provides a new improved support for better decision-making in the context of safety management. The scope can be one mode of transport but importantly also the whole transport system with its four transport modes. In an innovative way, the experienced events can also be brought into the same risk picture, showing a footprint of reality

next to the identified risks whose positions are based on estimations, judgments and assumptions. The different dimensions of severity can be considered in an integrated way making comparisons between different risks and intervention options possible. The need to ensure non-safety priorities (such as environmental sustainability) has been integrated in the process so that the outcome of the proposed process is not only addressing safety priorities but can address all the relevant priorities. There is also a specific provision for managing low-probability-high-impact risks through resilience building and maintaining awareness of uncertainty and possible surprises, instead of such risks dropping out of focus, which could be the case for classic risk management approaches.

The challenges of complex systems can never be fully compensated even with the best of methods. However, in the NRMF complexity has been identified as one of the key challenges from the beginning and a long list of techniques have been implemented for that purpose. The Cynefin framework is used to identify the type of system one is dealing with and several techniques adapted for complexity are used, like adaptive policies and experiments. Instead of one-shot actions, the developed approach proposes how a portfolio of experiments can be created and managed. The need to focus on the whole at the system level and the interactions between problems and between solutions are emphasized. Involving a diverse group of stakeholders is encouraged, helping in creating many different perspectives and in collecting knowledge. The proposed process includes the cycle of constant learning on the system and the success of the interactions. With all these conscious efforts, it can be expected that the NRMF handles complexity significantly better than existing methods.

As in today's world the remaining accidents in the very safe transport systems may have devastating consequences, it is hoped that this contribution helps in preventing them.

The area of application has been the transport safety agency. Implementation in another part of the transport system should be even easier – e.g. at an operator. There would be only one mode of transport, in a specific limited scope and the political and societal dimensions would be far more limited than for the agency. Due to the generic nature of the framework, it could be implemented also in a completely different field, with minor customization.

Future research

Perhaps the most fascinating area for future research are the adaptive interventions. How should experiments and adaptive policies be designed so that their combination creates the highest possible chances for success? The challenge of managing several portfolios of adaptive (and non-adaptive) interventions is also worth more research. There are contexts where experimenting may be particularly difficult, for instance due to the very high standardization of the activity and/or its criticality. A high-reliability environment may offer a thin flow of feedback on the success of an experiment: if pilots very rarely make a mistake in inputting data in the flight management computer with potentially catastrophic consequences, how could experiments be carried out with new procedures for data input or cross-checking in a way which would produce rapid feedback? Less standardized environments offer even easier opportunities for interesting research topics. For instance, the known risk-group of young drivers could be an easy target for a wide range of experiments which could be carried out in different towns or regions.

The work within the risk workshops is another interesting area. The topics include at least capturing and recording the necessary information and knowledge and being able to pass on the information to the decision group with enough detail, insight and context. The learning among the participants is another significant focus area. How can long-term learning be supported so that collective wisdom is generated? A key part of this is the capability to modellize/visualize at least parts of the system and to expose such models to collective scrutiny.

Third key area is the way the process could be supported with a dedicated software. Chapter 7.8 took the first steps in this direction by outlining some of the key features. It seems obvious that visual analytics could bring significant benefits to the process, both through visualization and by letting the user interact with the data in an iterative way.

In general, implementing the developed framework in several organizations would produce valuable feedback and enable further refinements and customization. Finally, it would be very interesting to adapt the framework to another domain outside transport and implement it there. The potential for a “universal risk currency” produced through the use of risk points suggests there could be an opportunity for novel research integrating or comparing risks (and risk perceptions) across different domains.

Personal publications

Nisula, J.M. 2014. From Safety Indicators to Measuring Risk – the Risk-Guided Transport Safety Agency. *Proceedings of the International Conference on Human-Computer Interaction in Aerospace 2014 (HCI-Aero)*; *Santa Clara, USA, July 30-August 1, 2014*.

Nisula, J.M. 2015. Modern approach for integrating safety events in a risk management process. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 3551-3559. London: Taylor & Francis Group.

Nisula, J.M. 2015. Creating an integrated risk picture for four modes of transport. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 3935-3943. London: Taylor & Francis Group.

Mazaheri A., Montewka J., Nisula J., Kujala P. 2015. Usability of accident and incident reports for evidence-based risk modeling – A case study on ship grounding reports. *Safety Science* 76: 202–214.

References

Aalto university 2012. *Marine technology annual report 2012*. Department of applied mechanics. School of engineering. Espoo: Aalto university.

Abrahamsen, H.B. & Abrahamsen, E.B. 2015. On the appropriateness of using the ALARP principle in safety management. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 773-777. London: Taylor & Francis Group.

Abrahamsen, E.B., Aven, T., Vinnem J.E., Wiencke, H.S. 2004. Safety management and the use of expected values. *Risk, Decision and Policy* 9(4): 347-357.

Ackoff, R. 1971. Towards a system of systems concepts. *Management science* 17(11): 661-671.

Ackoff, R. 1989. From data to wisdom. *Journal of applied systems analysis* 16: 3-9.

Ackoff, R. 1999. Re-creating the corporation. A design of organizations for the 21st century. New York: Oxford university press.

Ahteensuu, M. 2013. The precautionary principle and the justifiability of three imperatives. *Homo oeconomicus* 30(1): 17-36.

Airline Risk Management Solutions (ARMS) Working Group 2010. *Methodology for Operational Risk Assessment for Aviation Organizations*. Available at www.skybrary.aero

Allianz 2015. *Safety and Shipping review 2015*. Allianz Global Corporate & Specialty. Munich: Allianz.

Amalberti, R. 1999. Les effets pervers de l'ultrasécurité. *La recherche* 319: 66-70.

Amalberti, R. 2001. The paradoxes of almost totally safe transportation systems. *Safety Science* 37: 109-126.

Amalberti, R. 2013. *Navigating safety*. Dordrecht: Springer.

Anderson, P. 1999. Complexity theory and organization science. *Organization science* 10(3): 216-232.

Antonsen, S. 2009. Safety culture and the issue of power. *Safety Science* 47: 183-191.

Aven, T. 2007. A unified framework for risk and vulnerability analysis covering both safety and security. *Reliability engineering and system safety* 92: 745-754.

Aven, T. 2008. A semi-quantitative approach to risk analysis, as an alternative to QRAs. *Reliability engineering and system safety* 93: 768-775.

Aven, T. 2011a. On the new ISO guide on risk management terminology. *Reliability engineering and system safety* 96: 719-726.

Aven, T. 2011b. On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. *Risk analysis* 31(4): 515-522.

Aven, T. 2011c. On different types of uncertainties in the context of the precautionary principle. *Risk analysis* 31(10): 1515-1525.

Aven, T. 2012. The Risk concept - historical and recent development trends. *Reliability engineering and system safety* 99: 33-44.

Aven, T. 2013a. Practical implications of the new risk perspectives. *Reliability engineering and system safety* 115: 136-145.

Aven, T. 2013b. On how to deal with deep uncertainties in a risk assessment and management context. *Risk analysis* 33(12): 2082-2091.

Aven, T. 2013c. On the meaning and use of the risk appetite concept. *Risk analysis* 33(3): 462-468.

Aven, T. 2013d. On the meaning of a black swan in a risk context. *Safety Science* 57: 44-51.

Aven, T. 2014a. *Risk, surprises and black swans. Fundamental concepts in risk assessment and risk management*. Abingdon/New York: Routledge.

Aven, T. 2014b. What is safety science? *Safety Science* 67: 15-20.

Aven, T. 2015. Implications of black swans to the foundations and practice of risk assessment and management. *Reliability engineering and system safety* 134: 83-91.

Aven, T. & Krohn, B.S. 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability engineering and system safety* 121: 1-10.

Aven, T. & Reniers, G. 2013. How to define and interpret a probability in a risk and safety setting. *Safety Science* 51: 223-231

Aven, T. & Vinnem, J.E. 2005. On the use of risk acceptance criteria in the offshore oil and gas industry. *Reliability engineering and system safety* 90: 15-24.

Aven, T., Vinnem, J.E., Wiencke, H.S. 2007. A decision framework for risk management, with application to the offshore oil and gas industry. *Reliability engineering and system safety* 92: 433-448.

Aven, T. & Ylönen, M. 2016. Safety regulations: Implications of the new risk perspectives. *Reliability engineering and system safety* 149: 164-171.

Aven, T. & Zio, E. 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision-making. *Reliability engineering and system safety* 96: 64-74.

Balsa, A.I., Gandelman, N., González, N. 2015. Peer effects in risk aversion. *Risk analysis* 35(1): 27-43.

Beckerman, L.P. 2000. Application of complex systems science to systems engineering. Systems Engineering 3(2): 96-102.

Bergström, J., Dahlström, N., Van Winsen, R., Lützhöft, M., Dekker, S.W.A, Nyce, J. 2009. Rule- and role-retreat: an empirical study of procedures and resilience. Journal of maritime research 6(1): 75-90.

Bergström, J., Dahlström, N., Dekker, S.W.A. 2011. Training organizational resilience in escalating situations. In Hollnagel, Pariès, Woods, Wreathall (eds.) *Resilience engineering in practice*: 45-57. Farnham: Ashgate.

Berner, C.L., Flage, R. 2016. Comparing and integrating the NUSAP notational scheme with an uncertainty-based risk perspective. *Reliability engineering and system safety* 156: 185-194.

Beurden (van), E.K., Kia, A.M., Zask, A., Dietrich, U., Rose, L. 2013. Making sense in a complex landscape: how the Cynefin framework from Complex Adaptive Systems Theory can inform health promotion practice. Health promotion international 28(1): 73-83.

Bjerga, T. & Aven, T. 2015. Adaptive risk management using new risk perspectives – an example From the oil and gas industry. *Reliability engineering and system safety* 134: 75-82.

Bjerga, T., Aven, T., Zio, E. 2016. Uncertainty treatment in risk analysis of complex systems: The cases of STAMP and FRAM. *Reliability engineering and system safety* 156: 203-209.

Boeing commercial airplanes 2016. *Statistical summary of commercial jet airplane accidents. Worldwide operations, 1959-2015*. Seattle: Boeing.

Booth, J. 2012. *Transport Canada's risk based surveillance and planning system*. Presentation in European Aviation Safety Agency (EASA) conference "Safety Oversight – Managing safety in a performance based

regulatory environment". 10-11 October 2012, Cologne, Germany. Presentation available at: http://easa.europa.eu/conferences/pbo/ . Accessed 8-Feb-2016.

Carlsson, F., Johansson-Stenman, O., Martinsson, P. 2004. Is transport safety more valuable in the air? *The journal of risk and uncertainty* 28(2): 147-163.

Checkland, P.B. & Haynes, M.G. 1994. Varieties of systems thinking: the case of soft systems methodology. *System dynamics review* 10(2-3): 189-197.

Chilton, S., Covey, J., Hopkins, L., Jones-Lee, M., Loomes, G., Pidgeon, N., Spencer, A. 2002. Public perceptions of risk and preference-based values of safety. *The journal of risk and uncertainty* 25(3): 211-232.

Cilliers, P. 1998. *Complexity and postmodernism. Understanding complex systems*. London & New York: Routledge.

Civil Aviation Safety Authority (CASA, Australia) 2014. *Sector risk profile for the aerial application sector*. Canberra: CASA.

Columbia Accident Investigation Board (CAIB) 2003. *Columbia Accident Investigation Board report, volume I*. Washington D.C.: CAIB.

Cooke, D.L. & Rohleder, T.R. 2006. Learning from incidents: from normal accidents to high reliability. *System Dynamics Review* 22(3): 213–239.

Cox, L.A. 2012. Confronting deep uncertainties in risk analysis. *Risk Analysis* 32(10):1607–1629.

De Martino, B., Kumaran, D., Seymour, B., Dolan, R.J. 2006. Frames, biases and rational decision-making in the human brain. *Science* 313(5787): 684-687.

Dekker, S.W.A. 2003. Illusions of explanation: a critical essay on error classification. *International journal of aviation psychology* 13(2): 95-106.

Dekker, S.W.A. 2011. *Drift into failure. From hunting broken components to understanding complex systems*. Farnham: Ashgate.

Dekker, S.W.A., Cilliers, P., Hofmeyr, J-H. 2011. The complexity of failure: implications of complexity theory for safety investigations. *Safety science* 49: 939-945.

Dekker, S.W.A. 2014. The bureaucratization of safety. *Safety Science* 70: 348-357

Delorme, R., Lassarre, S. 2014. A new theory of complexity for safety research. The case of the long-lasting gap in road safety outcomes between France and Great Britain. *Safety science* 70: 488-503.

Duijm, N.J. 2015. Recommendations on the use and design of risk matrices. *Safety Science* 76: 21-31.

Ersdal, G. & Aven, T. 2008. Safety regulations: Risk informed decision-making and its ethical basis. *Reliability engineering and system safety* 93: 197-205.

Eurocontrol 2013. *From Safety-I to Safety-II: A white paper*. Brussels: Eurocontrol.

Eurocontrol 2014. *Systems Thinking for Safety: Ten Principles. A White Paper. Moving towards Safety-II*. Brussels: Eurocontrol.

European Aviation Safety Agency (EASA) 2016. EASA – Looking forward. *Presentation given by Certification Director Trevor Woods, 9 February 2016*. Available at: https://www.trafi.fi/filebank/a/1455105666/ecec7fb73d57c8b402065ff63d053325/19774-Trevor_Woods_-_TWO_Helsinki_9_Feb_2016_Final.pdf . Accessed 2-Mar-2017.

European Co-ordination Centre for Aviation Incident Reporting Systems (ECCAIRS) 2010. *ECCAIRS 4.2.8 Data Definition Standard, Explanatory Factors*. Ispra: Joint Research Centre (JRC) of the European Union.

European general aviation safety strategy working group 2012. *European general aviation safety strategy*. Available at: https://www.easa.europa.eu/system/files/dfu/European%20GA%20Safety%20Strategy.pdf . Accessed 2-Mar-2017.

European Rail Research Advisory Council (ERRAC) 2016. *Regional and suburban railways – market analysis update. Brussels*: ERRAC.

European Transport Safety Council (ETSC) 2015. *Ranking EU progress on road safety. 9th Road safety performance index report*. Brussels: ETSC.

European Union Agency for Railways (ERA) 2009. *Guide for the application of the Commission Regulation on the adoption of a common safety method on risk evaluation and assessment as referred to in Article 6(3)(a) of the Railway Safety Directive*. ERA/GUI/01-2008/SAF. Valenciennes: ERA.

European Union Agency for Railways (ERA) 2015a. *SMS – Legal basis*. Available at: http://www.era.europa.eu/tools/sms/Pages/Legal-basis-.aspx . Accessed 20-Jan-2017.

European Union Agency for Railways (ERA) 2015b. *The SMS Wheel*. Available at: http://www.era.europa.eu/tools/sms/Pages/SMS.aspx#SMS_Wheel . Accessed 20-Jan-2017.

European Union Agency for Railways (ERA) 2016. *Railway safety performance in the European Union 2016*. Biennial report. Valenciennes: ERA.

Fairbanks, R.J., Wears, R.L., Woods, D.D., Hollnagel, E., Plsek, P., Cook, R.I. 2014. Resilience and resilience engineering in health care. The joint commission journal on quality and patient safety 40(8): 376-383.

Federal Aviation Administration (FAA) 1988. System design and analysis. *Advisory circular 25.1309-1A, dated 21 June 1988*. Washington D.C.:FAA.

Federal Aviation Administration (FAA) 2011. System safety analysis and assessment for part 23 airplanes. *Advisory circular 23.1309-1E, dated 17 November 2011*. Washington D.C.:FAA.

Federal Aviation Administration (FAA) 2014. Air Traffic Organization. *Safety Management System Manual*. Version 4.0. Washington D.C.:FAA.

Feduzi, A., Runde, J. 2014. Uncovering unknown unknowns: towards a Baconian approach to management decision-making. *Organizational behavior and human decision processes* 124: 268-283.

Fischoff, B. 1975. Hindsight≠foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of experimental psychology: human perception and performance* 1(3): 288-299.

Flage, R., Aven, T. 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). *Reliability and Risk Analysis: Theory and Applications* 2(13): 9–18.

Flage, R., Aven, T. 2015. Emerging risk – conceptual definition and a relation to black swan type of events. *Reliability engineering and system safety* 144: 61-67.

Flood, R.L. 2010. The relationship of 'systems thinking' to action research. *Systemic practice and action research* 23: 269-284.

Forrester, J.W. 1971. Testimony for the subcommittee on urban growth of the committee on banking and currency, U.S. house of representatives, Oct 7, 1970. *Technology Review*. Massachusetts: Alumni association of the Massachusetts Institute of Technology.

Forrester, J.W. 1994. System dynamics, systems thinking, and soft OR. *System dynamics review* 10(2-3): 245-256.

French, S. 2012. Cynefin, statistics and decision analysis. *Journal of the operational research society* 64(4): 547-561.

French, S. 2015. Cynefin: uncertainty, small worlds and scenarios. *Journal of the operational research society* 66: 1635-1645.

Funtowicz, S.O., Ravetz, J.R. 1990. *Uncertainty and quality in science for policy*. Berlin: Springer science & business media.

Gershenson, C. 2013. The implications of interactions for science and philosophy. *Foundations of Science* 18(4), 781-790.

Gilbert, S.F., Sarkar, S. 2000. Embracing complexity: organicism for the 21st century. *Developmental dynamics* 219: 1-9.

Gorzen-Mitka, I., Okreglicka, M. 2014. Improving decision making in complexity environment. *Procedia economics and finance* 16: 402-409.

Grote, G. 2015. Promoting safety by increasing uncertainty – Implications for risk management. *Safety science* 71: 71-79.

Hafver, A., Lindberg, D.V., Jakopanec, I., Pedersen, F.B., Flage, R., Aven, T. 2015. Risk – from concept to decision making. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 779-784. London: Taylor & Francis Group.

Hale, A. 2001. Conditions of occurrence of major and minor accidents. *Journal of the institution of occupational safety and health* 5, issue 1: 7-21.

Hasan, H., Kazlauskas, A. 2009. Making sense of IS with the Cynefin framework. *Proceedings of the Pacific Asia Conference on Information Systems (PACIS)*: 1-13. Hyderabad: Indian School of Business.

Higgins, D. 2015. Black swan events and property asset management: Redefining place and space on global organisations property decisions. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 2755-2761. London: Taylor & Francis Group.

Hokstad, P., Vatn, J., Aven, T. and Sörum, M., 2004. Use of risk acceptance criteria in Norwegian offshore industry: Dilemmas and challenges. *Risk, decision and policy* 9(3):193-206.

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

Hollnagel, E. 2011. RAG – The resilience analysis grid. In Hollnagel, Pariès, Woods, Wreathall (eds.) *Resilience engineering in practice*: 275-296. Farnham: Ashgate.

Hollnagel, E. 2012. *FRAM: The functional resonance analysis method*. Farnham: Ashgate.

Hollnagel, E. 2014. *Safety-I and Safety-II*. Farnham: Ashgate.

Hollnagel, E., Woods, D.D., Leveson, N. 2006. *Resilience engineering. Concepts and precepts*. Aldershot: Ashgate.

Hänninen, M. & Kujala, P. 2012. Influences of variables on ship collision probability in a Bayesian belief network model. *Reliability Engineering and System Safety* 102: 27–40.

Hänninen, M., Sladojevic, M., Tirunagari, S., Kujala, P. 2013. Feasibility of collision and grounding data for probabilistic accident modeling. In Amdahl, Ehlers & Leira (Eds) *Collision and Grounding of Ships and Offshore Structures*. London: Taylor & Francis Group.

Inayatullah, S. 2002. Reductionism or layered complexity? The futures of futures studies. *Futures* 34: 295-302.

Insua, D.R., Alfaro, C., Gomez, J., Hernandez-Coronado, P., Bernal, F. (in press). A framework for risk management decisions in aviation safety at state level. *Reliability Engineering & System Safety*.

International Air Transport Association (IATA) 2013. *Evidence-Based Training implementation guide*. Montreal: IATA.

International Air Transport Association (IATA) 2014. *Data report for Evidence-Based Training*. Montreal: IATA.

International Association of Public Transport (UITP) 2016. *Local public transport in the European Union – Statistics brief*. Version 2, September 2016. Brussels: UITP.

International Civil Aviation Organization (ICAO) 2001. *Annex 13 to the convention on international civil aviation: Aircraft accident and incident investigation.* Ninth edition, July 2001. Montreal: ICAO

International Civil Aviation Organization (ICAO) 2009. Safety Management Systems (SMS) Course. Module 5 – Risks. Revision 13. Powerpoint presentation. Montreal: ICAO

International Civil Aviation Organization (ICAO) 2010. *High-level safety conference 2010, working paper.* Montreal: ICAO

International Civil Aviation Organization (ICAO) 2013. *Safety Management Manual.* Doc 9859. Third edition. Montreal: ICAO

International Civil Aviation Organization (ICAO) 2015. *Annual report of the council - 2014.* Appendix 1. Montreal: ICAO

International Civil Aviation Organization (ICAO) 2016. *Global air navigation plan 2016-2030*. Montreal: ICAO

International Civil Aviation Organization (ICAO) 2017. *SARPs – Standards and Recommended practices. Initial introduction of ICAO Safety Management SARPs.* Available at: http://www.icao.int/safety/SafetyManagement/Pages/SARPs.aspx . Accessed 20-Jan-2017.

International Maritime Organization (IMO) 2002. *Guidelines for formal safety assessment (FSA) for use in the IMO rule-making process (MSC/Circ.1023 - MEPC/Circ.392)*. London: IMO.

International Maritime Organization (IMO) 2006. *Amendments to the guidelines for formal safety assessment (FSA) for use in the IMO rulemaking process (MSC/Circ.1023 - MEPC/Circ.392)*. London: IMO.

International Organization for Standardization (ISO) 2009. *ISO 31000 Risk Management – Principles and Guidelines*. Geneva: ISO.

International Transport Forum 2015. *Road safety annual report 2015*. Paris: International Transport Forum.

Jackson, M.C. 1994. Critical systems thinking: beyond the fragments. *System dynamics review* 10(2-3): 213-229.

Jamroz, K., Kadziński, A., Chruzik, K., Szymanek, A., Gucma, L., & Skorupski, J. 2010. TRANS-RISK-an integrated method for risk management in transport. *Journal of KONBiN* 13(1), 209-220.

Johansen, I.L & Rausand, M. 2014. Foundations and choice of risk metrics. *Safety Science* 62: 386-399.

Johnson, C.W. 2003. A handbook of incident and accident reporting. Glasgow: Glasgow university press. Available at: http://www.dcs.gla.ac.uk/~johnson/book/ . Accessed 15-Jan-2017.

Johnson, C.W. 2007. The paradoxes of military risk assessment: will the enterprise risk assessment model, composite risk management and associated techniques provide the predicted benefits? In *proceedings of the 25th international systems safety conference*, Baltimore, USA: 859-69. Unionville, VA, USA: International systems safety society.

Johnson, C.W., Shea, C. 2007. A comparison of the role of degraded modes of operation in the causes of accidents in rail and air traffic management. *Proceedings of the 2nd Institution of Engineering and Technology international conference on system safety*: 89-94. London, 22-24 October 2007. London: Institution of engineering and technology (IET).

Jones-Lee, M., Aven, T. 2011. ALARP – what does it really mean? *Reliability engineering and system safety* 96: 877-882.

Jonkman, S.N., van Gelder, P.H.A.J.M., Vrijling, J.K. 2003. An overview of quantitative risk measures for loss of life and economic damage. Journal of hazardous materials A99: 1-30.

Jore, S.H., Egeli, A. 2015. Risk management methodology for protecting against malicious acts — are probabilities adequate means for describing terrorism and other security risks?. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 807-815. London: Taylor & Francis Group.

Kahneman, D., Tversky, A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47(2): 263-292.

Khan, F., Rathnayaka, S., Ahmed, S. 2015. Methods and models in process safety and risk management: Past, present and future. *Process safety and environmental protection* 98: 116-147.

Kontogiannis, T., Leva, M.C., Balfe, N. (in press). Total safety management: principles, processes and methods. *Safety science*.

Kurtz, C.F., Snowden, D.J. 2003. The new dynamics of strategy: sense-making in a complex and complicated world. *IBM systems journal* 42(3): 462-483.

Ladan, M. & Hänninen, M. 2012. *Data Sources for Quantitative Marine Traffic Accident Modeling*. Espoo: Aalto University.

La Porte, T.M. 2006. Organizational strategies for complex systems, Resilience, Reliability, and Adaptation. In Auerswald, Branscomb, La Porte, Michel-Kerjan (eds.) *Seeds of Disaster, Roots of Response*: 135-154. Cambridge: Cambridge University Press.

Laprie, J. C. 2008. From dependability to resilience. In *38th IEEE/IFIP Int. conference on dependable systems and networks*: G8-G9. New York: IEEE.

Larouzée, J. & Guarnieri, F. 2015. From theory to practice: Itinerary of Reasons' Swiss Cheese Model. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 817-824. London: Taylor & Francis Group.

Lee, R. & De Landre, J. 2000. *The Systemic Incident Analysis Model (SIAM)*. Canberra: Australian Transport Safety Bureau.

Le Coze, J-C. 2005. Are organizations too complex to be integrated in technical risk assessment and current safety auditing? *Safety Science* 43: 613-638.

Leveson, N., Dulac, N., Zipkin, D., Cutcher-Gershenfeld, J., Carroll, J., Barrett, B. 2006. Engineering resilience into safety-critical systems. In Hollnagel et al. (eds) *Resilience engineering. Concepts and precepts*: 95-123. Aldershot: Ashgate.

Leveson, N.G. 2011. Applying systems thinking to analyze and learn from events. *Safety Science* 49: 55-64.

Leviäkangas, P., Aapaoja, A. 2015. Resilience of transport infrastructure systems. CSID journal of sustainable infrastructure development 1: 77-87.

Linkov, I., Satterstrom, F.K., Kiker, G., Batchelor, C., Bridges, T., Ferguson, E. 2006. From comparative risk assessment to multi-criteria decision analysis and adaptive management: Recent developments and applications. *Environment international* 32: 1072-1093.

Makridakis, S., Hogart, R.M., Gaba, A. 2009. Forecasting and uncertainty in the economic and business world. *International Journal of forecasting* 25: 794-812.

Makridakis, S. & Taleb, N.N. 2009. Decision-making and planning under low levels of predictability. *International Journal of forecasting* 25: 716-733.

Marais, K., Saleh, J.H., Leveson, N.G. 2005. Archetypes for organizational safety. *Safety Science* 44: 565-582.

Marchau, V.A.W.J., Walker, W.E., van Wee, G.P. 2010. Dynamic adaptive transport policies for handling deep uncertainty. *Technological forecasting & social change* 77: 940-950.

Marion, R., Uhl-Bien, M. 2001. Leadership in complex organizations. *The leadership quarterly* 12(4): 389-418.

Maurino, D.E., Reason, J., Johnston, N., Lee, R.B. 1995. *Beyond Aviation Human Factors*. Aldershot: Ashgate.

Meadows, D.H. 2008. *Thinking in systems*. White river junction (VT): Chelsea green publishing.

Meadows, D.H., Richardson, J., Bruckmann, G. 1982. *Groping in the dark. The first decade of global modelling*. Chichester: John Wiley & sons.

Mikulecky, D.C. 2001. The emergence of complexity: science coming of age or science growing old? Computers and chemistry 25: 341-348.

Montewka, J., Goerlandt, F., Kujala, P. 2014. On a systematic perspective on risk for formal safety assessment (FSA). *Reliability engineering and system safety* 127: 77-85.

Morel, G., Amalberti, R., Chauvin, C. 2008. Articulating the differences between safety and resilience: The decision-making process of professional sea-fishing skippers. *Human Factors* 50(1): 1-16.

Mosleh, A., Bier, V.M., Apostolakis, G. 1988. A critique of current practice for the use of expert opinions in probabilistic risk assessment. *Reliability engineering and system safety* 20(1): 63-85.

Moxnes, E. 2000. Not only the tragedy of the commons: misperceptions of feedback and policies for sustainable development. *System dynamics review* 16(4): 325-348.

National Patient Safety Agency (NPSA) 2008. *A risk matrix for risk managers*. London: NPSA.

Naweed, A., Rainbird, S., Dance, C. 2015. Are you fit to continue? Approaching rail systems thinking at the cusp of safety and the apex of performance. *Safety Science* 76: 101-110.

Nickerson, R., S. 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Review of general psychology* 2(2): 175-220.

Nisula, J.M. 2014a. From Safety Indicators to Measuring Risk – the Risk-Guided Transport Safety Agency. *Proceedings of the International Conference on Human-Computer Interaction in Aerospace 2014 (HCI-Aero)*; *Santa Clara, USA, July 30-August 1, 2014*.

Nisula, J.M 2014b. *Safety Factors in the 'Tiedosta Toimenpiteisiin' (TiTo) project*. Available at: http://www.trafi.fi/tietopalvelut/tutkimus_ja_kehittaminen/tiedosta_toimenpiteisiin. Helsinki: Finnish Transport Safety Agency (Trafi).

Nisula, J.M. 2015a. Modern approach for integrating safety events in a risk management process. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 3551-3559. London: Taylor & Francis Group.

Nisula, J.M. 2015b. Creating an integrated risk picture for four modes of transport. In Podofillini et al. (eds) *Safety and Reliability of Complex Engineered Systems*: 3935-3943. London: Taylor & Francis Group.

Paté-Cornell, E. 2002. Risk and uncertainty analysis in government safety decisions. *Risk analysis* 22(3): 633-646.

Paté-Cornell, E. 2012. On "black swans" and "perfect storms": risk analysis and management when statistics are not enough. *Risk analysis* 32(11): 1823-1833.

Paté-Cornell, E. & Cox, L.A.Jr. 2014. Improving risk management: from lame excuses to principled practice. *Risk analysis* 34(7): 1228-1239.

Perrow, C. 1984. *Normal accidents. Living with high-risk technologies*. USA: Basic books.

Persson, U., Norinder, A., Hjalte, K., Gralén, K. 2001. The value of statistical life in transport: findings from a new contingent valuation study in Sweden. *The journal of risk and uncertainty* 23(2): 121-134.

Queensland trucking association (QTA) 2009. A safety management system for small transport businesses. Stones corner: QTA. Available at: http://www.qta.com.au/resources/Documents/WHS/QTA%20SMS%20-%20V9%20FINAL%20-%20Full%20Version.pdf . Accessed 20-Jan-2017.

Rabin, M. & Thaler, R.H. 2001. Anomalies: risk aversion. *The journal of economic perspectives* 15(1): 219-232.

Rail Safety and Standards Board (RSSB) Limited 2009. *Guidance on the preparation and use of company risk assessment profiles for transport operators*. Issue 1. London: RSSB.

Rail Safety and Standards Board (RSSB) Limited 2014a. Risk based decision making. Developing continuous improvement. *Presentation by Colin Dennis (technical director, RSSB) at the EU railway conference*, Lille, 8-May-2014. London: RSSB.

Rail Safety and Standards Board (RSSB) Limited 2014b. *Safety risk model: risk profile bulletin, version 8.1*. London: RSSB.

Rankin, A., Lundberg, J., Woltjer, R., Rollenhagen, C., Hollnagel, E. 2014. Resilience in everyday operations: a framework for analyzing adaptations in high-risk work. *Journal of cognitive engineering and decision making* 8(1): 78–97.

Rasmussen, J., Svedung, I. 2000. Proactive risk management in a dynamic society. Karlstad: Swedish rescue services agency.

Reason, J. 1990. *Human Error*. Cambridge: Cambridge University Press.

Reason, J.1997. *Managing the risks of organizational accidents*. Aldershot: Ashgate.

Reason, J. 2000. Human error: models and management. *BMJ* 320: 768-770.

Reiman, T., Rollenhagen, C., Pietikäinen, E., Heikkilä, J. 2015. Principles of adaptive management in complex safety critical organizations. *Safety Science* 71: 80–92.

Rheinberger, C.M. 2010. Experimental evidence against the paradigm of mortality risk aversion. *Risk analysis* 30(4): 590-604.

Rowley, J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science* 33(2): 163-180.

Ruhl, J.B. 1996. Complexity theory as a paradigm for the dynamical law-and-society system: a wake-up call for legal reductionism and the modern administrative state. *Duke law journal* 45(5): 849-928.

Safety Management International Collaboration Group (SMICG) 2013. *Risk Based Decision Making Principles*. SMICG. Available at: http://www.skybrary.aero/index.php/Risk_Based_Decision_Making_Principles

Sagan, S., D. 2004. Learning from normal accidents. *Organization & environment* 17(1): 15-19.

Senge, P.M. 2006. *The fifth discipline. The art & practice of the learning organization*. Revised edition of 2006. London: Random house group.

Shyur, H-J. 2008. A quantitative model for aviation safety risk assessment. *Computers & Industrial engineering* 54: 34-44.

Sklet, S. 2004. Comparison on some selected methods for accident investigation. *Journal of hazardous materials* 111: 29-37.

Sluijs (van der), J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J., Risbey, J. 2005. Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. *Risk analysis* 25(2): 481-492.

Snowden, D.J. 2005a. From atomism to networks in social systems. *The learning organization* 12(6): 552-562.

Snowden, D.J. 2005b. Multi-ontology sense making: a new simplicity in decision making. *Informatics in primary care* 13: 45-53.

Snowden, D.J. 2005c. Strategy in the context of uncertainty. *Handbook of business strategy* 6(1): 47-54.

Snowden, D.J., Boone, M.E. 2007. A leader's framework for decision making. *Harvard business review*, Nov 2007. Boston, MA: Harvard Business School Publishing.

Society for Risk Analysis (SRA) 2015a. SRA Glossary. Available at: http://sra.org/sites/default/files/pdf/SRA-glossary-approved22june2015-x.pdf

Society for Risk Analysis (SRA) 2015b. Risk Analysis Foundations. Available at: http://www.sra.org/sites/default/files/pdf/FoundationsMay7-2015-sent-x.pdf

Sornette, D., Ouillon, G. 2012. Dragon-kings: mechanisms, statistical methods and empirical evidence. *The European physical journal special topics* 205: 1-26.

Standards Australia 2009. *Risk Management – Principles and Guidelines. AS/NZS ISO 31000:2009*. Sydney/Wellington: Standards Australia/Standards New Zealand.

Statistics Finland 2016. *Statistics on road traffic accidents 2015*. Transport and Tourism statistics, preliminary data dated 15-Jun-2016. Helsinki: Statistics Finland. Available at: http://tilastokeskus.fi/til/ton/2015/ton_2015_2016-06-15_en.pdf . Accessed 21-Oct-2016.

Sterman, J.D. 1994. Learning in and about complex systems. *System dynamics review* 10(2-3): 291-330.

Stroeve, S.H., Blom, H.A.P., Bakker, G.J. 2013. Contrasting safety assessments of a runway incursion scenario: Event sequence analysis versus multi-agent dynamic risk modelling. *Reliability Engineering and System Safety 109*: 133–149.

Sturmberg, J.P., Martin, C.M. 2008. Knowing – in medicine. *Journal of evaluation in clinical practice* 14: 767–770.

Sutcliffe, K.M. 2011. High reliability organizations (HROs). *Best practice & Research clinical anaesthesiology* 25: 133-144.

Swanson, D., Barg, S., Tyler, S., Venema, H., Tomar, S., Bhadwal, S., Nair, S., Roy, D., Drexhage, J. 2010. Seven tools for creating adaptive policies. *Technological forecasting & social change* 77: 924-939.

Taig, T., Hunt, M. 2012. Review of LU and RSSB safety risk models. A report produced for the office of rail regulation. Issue 01. Northwich: TTAC Ltd.

Taleb, N.N. 2010. *The Black Swan*. 2nd edition. New York: Random House.

Technical Research Centre of Finland (VTT) 2017. *Trafin riskianalyysimenetelmän arviointi* (assessment of the risk analysis method of Trafi). Document VTT-CR-01046-17 dated 8th March 2017. Espoo: VTT.

Thekdi, S., Aven, T. 2016. An enhanced data-analytic framework for integrating risk management and performance management. *Reliability Engineering and System Safety 156*: 277-287.

Tickner, J., Kriebel, D. 2006. The role of science and precaution in environmental and public health policy. In Fisher, E., Jones, J., von Schomberg, R. (eds.) *Implementing the precautionary principle*. Northhampton, MA, USA: Edward Elgar publishing.

Trafi (Finnish Transport Safety Agency) 2012. *Strategia 2020*. Trafi publications 22/2012 (in Finnish). Helsinki: Trafi.

Trafi (Finnish Transport Safety Agency) 2013a. *Finnish annual maritime safety review 2013*. Helsinki: Trafi.

Trafi (Finnish Transport Safety Agency) 2013b. *Finnish annual aviation safety review 2013*. Helsinki: Trafi.

Trafi (Finnish Transport Safety Agency) 2014a. *Liikenteen Turvallisuusvirasto Trafi*. General presentation about Trafi (in Finnish). Helsinki: Trafi.

Trafi (Finnish Transport Safety Agency) 2014b. *Finnish railways 2014 – safety and environmental impacts*. Helsinki: Trafi.

Trafi (Finnish Transport Safety Agency) 2015a. *Suomen ilmailun tila 2015*. Trafi homepage. http://katsaukset.trafi.fi/media/katsaukset/ilmailu/suomen-ilmailun-tila-2015.pdf. Accessed 9-Feb-2016.

Trafi (Finnish Transport Safety Agency) 2015b. *About Trafi*. Trafi homepage. URL:http://www.trafi.fi/en/about_trafi. Accessed 28-Sep-2015.

Trafi (Finnish Transport Safety Agency) 2017. *Suomen ilmailun turvallisuusohjelma 2017* (Finnish Aviation Safety Programme 2017). Helsinki: Trafi.

Transportstyrelsen (Swedish Transport Agency) 2014. *Riktlinje för riskhantering och riskbaserad tillsyn*. (guidelines for risk management and risk-based treatment) Document TSG 2014-1394 dated 15-Sep-2014. Norrköping: Transportstyrelsen.

Tversky, A. & Kahneman, D. 1973. Availability: a heuristic for judging frequency and probability. *Cognitive psychology* 5(2): 207-232.

United Nations 2015. Transforming our world: the 2030 Agenda for Sustainable Development. *Resolution 70/1 adopted by the General Assembly on 25 September 2015*. New York: United Nations.

United Nations 2016. Economic and Social Council, Economic Commission for Europe, Inland Transport Committee. Working party on transport trends and economics. *Transport Trends and Economics 2016–2017: Achievement of Sustainable Development Goals through the development of Sustainable Transport*. Twenty-ninth session, Geneva, 5-7 September 2016. Document ECE/TRANS/WP.5/2016/5. Geneva: United Nations.

United Nations Conference on Trade and Development (UNCTAD) 2015. *Review of maritime transport 2015*. United Nations publication E.15.II.D.6. Geneva: United Nations

Urry, J. 2005. The complexity turn. *Theory, Culture & Society* 22(5): 1–14.

Vanem, E. 2012. Ethics and fundamental principles of risk acceptance criteria. *Safety Science* 50: 958-967.

Villa, V., Paltrinieri, N., Khan, F., Cozzani, V. 2016. Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Safety Science* 89: 77-93.

Viscusi, W.K., Aldy, J.E., 2003. The value of statistical life: a critical review of market estimates throughout the world. *The journal of risk and uncertainty* 27(1): 5-76.

Walker, W.E., Marchau, V.A.W.J., Swanson, D. 2010. Addressing deep uncertainty using adaptive policies: Introduction to section 2. *Technological Forecasting and Social Change* 77(6): 917–923.

Waring, A. 2015. Managerial and non-technical factors in the development of human-created disasters: a review and research agenda. Safety science 79: 254-267.

Wilson, R.S., Arvai, J.L., Arkes, H.R. 2008. My loss is your loss . . . sometimes: loss aversion and the effect of motivational biases. *Risk analysis* 28(4): 929-938.

Wolff, J. 2005. *Railway safety and the ethics of the tolerability of risk*. Study commissioned by the Rail Safety and Standards Board (RSSB). London: RSSB.

Woods, D.D., Johannesen, L.J., Cook, R.I, Sarter, N.B. 1994. Behind Human Error: Cognitive systems, computers and hindsight. Crew Systems Ergonomics Information Analysis Center (CSERIAC) State-of-the-art report 94-01. Ohio: CSERIAC.

Woods, D.D., Cook, R.I 2002. Nine steps to move forward from error. *Cognition, technology & work* 4: 137-144.

Woods, D.D. 2003. Creating foresight: How resilience engineering can transform NASA's approach to risky decision making. *Testimony on the future of NASA for committee on commerce, science and transportation, John McCain, Chair, October 29, 2003*.

Woods, D.D., Cook, R.I 2006. Incidents – Markers of resilience or brittleness? In Hollnagel, Woods, Leveson (eds.) *Resilience Engineering. Concepts and Precepts*: 69-76. Aldershot: Ashgate.

Woods, D.D. 2014. The future seems implausible, the past incredible. *Presentation at the Eurocontrol System Safety & Human Performance conference*, Lisbon, 24-26-Sep-2014.

Woods, D.D. 2015. Four concepts for resilience and the implications for the future of resilience engineering. *Reliability engineering & system safety* 141: 5-9.

World Heath Organization (WHO) 2015. *Global status report on road safety 2015*. Geneva: WHO.

Wreathall, J. 2006. Properties of resilient organizations: an initial view. In Hollnagel, Woods, Leveson (eds.) *Resilience engineering. Concepts and precepts*. Aldershot: Ashgate.

Wren, J., Barrell, K. 2010. The costs of injury in New Zealand and methods for prioritising resource allocation. *A background briefing paper to inform the evaluation of the New Zealand injury prevention strategy*. Wellington, New Zealand: New Zealand Injury Prevention Secretariat.

Abstract

Over the last ten-fifteen years, science has made significant advances in fields relevant for risk management. However, current risk management practices in industry have not yet benefitted from these developments. The research question addressed in this dissertation is: What kind of risk management framework should be used for managing transport risks when the modern risk perspectives and the latest understanding of safety are embraced, and the transport system is considered a complex adaptive system?

The focus of this research is on transport risks, taking the perspective of a national transport safety agency, tasked with overseeing safety across several modes of transport, including aviation, maritime, railway and road safety.

The scientific literature on risk and risk assessment, safety and safety management, as well as complex adaptive systems are reviewed. The research illustrates that a modern risk perspective recognizes the importance of uncertainty and strength of knowledge in risk analysis, as well as the role of surprises. The transport system is identified as a complex adaptive system, characterized by a high number of interactions, emergence, multiple feedback loops, nonlinear phenomena, unpredictability and counter-intuitiveness. The recommended ways to interact with such complex systems and to try to achieve positive change are explained. Concepts related to safety management are also investigated, especially the concept of resilience, which is interpreted as graceful extensibility of teams or organizations, or as sustained adaptability. Evidence of existing risk management frameworks in both the industry and scientific literature is outlined and reference is made to the international ISO 31000 standard for risk management. Based on the literature review, a set of criteria for a modern risk management process is developed.

A risk management framework for managing transport risks which embraces modern risk perspectives and accounts for the transport system as a complex adaptive system is proposed. It enables risks in all transport modes to be presented in a single risk picture and supports decision-making to maximize the safety impact achievable with limited resources. The impact is further enhanced by intervention strategies such as adaptive policies and experimentation, which are well-suited to complex systems. The framework is validated against the criteria developed, and by comparison to existing methods. A case study presents the on-going implementation of the developed risk management framework at the Finnish Transport Safety Agency. Both the proposed risk management framework and the dissertation are structured according to the ISO 31000 framework.

Keywords: Risk Management, Risk Management Framework, Complex Adaptive System, Transport Safety, ISO 31000

Résumé

La science a connu des avancées significatives en matière de gestion du risque au cours de la dernière décennie. Toutefois, les pratiques actuelles de gestion du risque dans le domaine industriel n'ont pas tiré tout le profit de ces développements. Le sujet de recherche de cette thèse peut être formulé ainsi : comment bâtir un cadre de gestion du risque afin de gérer les risques dans le transport, en adoptant les perspectives modernes du risque et les dernières connaissances de sécurité, tout en considérant le système de transport comme un système adaptatif complexe ? Ceci, à travers la perspective d'une agence nationale de la sécurité des transports, dont la responsabilité est la supervision de la sécurité de plusieurs modes de transport, incluant l'aérien, le maritime, le ferroviaire et le routier.

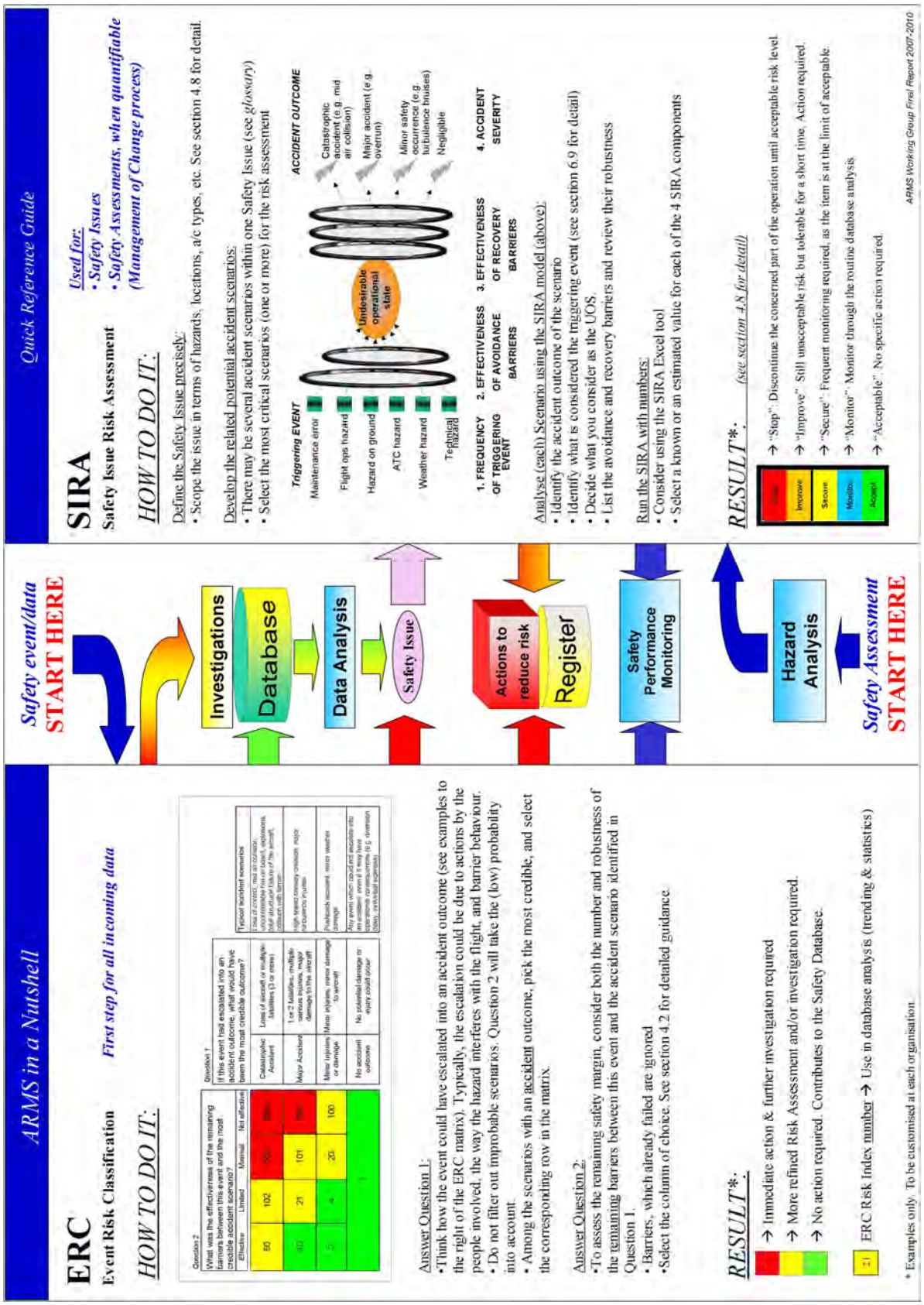
La connaissance scientifique actuelle est passée en revue pour les sujets de risques et d'appréciation du risque, de sécurité et de gestion de la sécurité ainsi que les systèmes adaptatifs complexes. L'approche moderne du risque implique reconnaitre l'importance de l'incertitude et la solidité des connaissances dans l'analyse du risque ainsi que le rôle des imprévus. Le système de transport est identifié comme un système adaptatif complexe. De tels systèmes se caractérisent par un large volume d'interactions, de nombreuses boucles de rétroaction, des phénomènes non-linéaires, l'émergence, l'imprévisibilité et la contre-intuitivité. Sont étudiées les façons recommandées d'interagir avec les systèmes complexes afin de tenter de parvenir à un changement positif. Les concepts relatifs à la gestion de la sécurité sont également présentés et notamment le concept de résilience qui peut être interprété soit comme une élégante extensibilité des équipes ou des organisations, soit comme une adaptabilité continue. Les cadres existants de management du risque sont revus à la fois dans l'industrie et dans la littérature scientifique ainsi que la norme internationale ISO 31000. Basé sur l'état de l'art, un ensemble de critères pour un processus moderne de management du risque est développé.

Le cadre proposé de gestion du risque dans le transport comprend des perspectives modernes du risque et considère le système de transport comme un système adaptatif complexe. Il permet de présenter les risques des différents modes de transport dans une visualisation globale de risque et de l'utiliser en tant que support pour prise de décision afin d'optimiser l'impact sur la sécurité avec les ressources qui sont toujours limitées. L'impact est encore renforcé par les moyens d'intervention tels que les stratégies adaptives et l'expérimentation, qui sont bien adaptés aux systèmes complexes. Elle est validée selon les critères élaborés et par comparaison avec les cadres existants. Le cadre proposé de gestion du risque ainsi que la thèse sont tous deux structurés d'après la norme ISO 31000. Enfin une étude de cas présente la mise en œuvre actuelle de cette nouvelle approche à l'Agence Nationale Finlandaise de la Sécurité des Transports.

Mots-clés: Gestion du risque, Management du risque, Système adaptatif complexe, Transport, Sureté de Fonctionnement, ISO 31000

Appendices

Appendix 1: ARMS process (the so-called "quick reference guide")

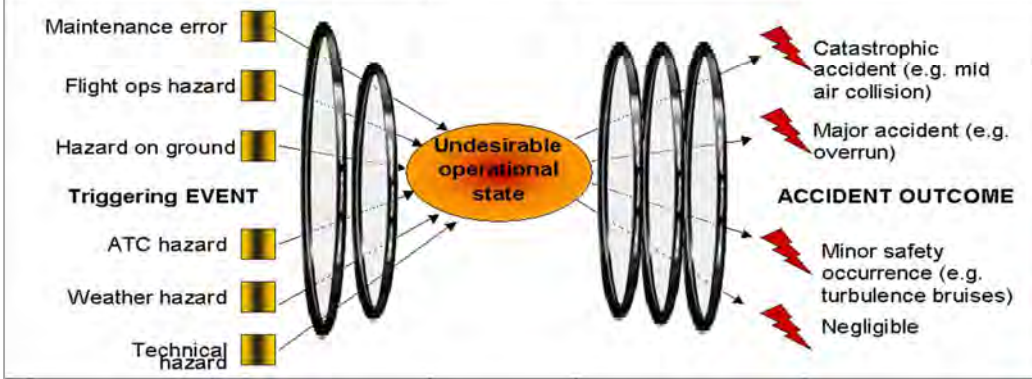


Appendix 2: ARMS Safety Issue Risk Assessment (SIRA) example excel tool

SAFETY ISSUE RISK ASSESSMENT (SIRA) TOOL

Fill in the grey areas!

1	**Safety Issue title:**				
2	**Define/scope the SI:**				
	Description of Hazard(s)				
	Description of Scenario				
	A/C types				
	Locations				
	Time period under study				
	Other				
3	**Analysis of potential Accident Scenario**				
	3.1 Triggering event		3.2 Undesirable Operational State		3.3 Accident Outcome



4	**Describe the barriers**				
		4.1 To avoid the UOS		4.2 To recover before the Accident	
5	**Risk Assessment**				
	The estimated frequency of the triggering event (per flight sectors) is:	The barriers will **fail** in AVOIDING the UOS...		The barriers will **fail** in RECOVERING the situation before the ACCIDENT...	The accident severity would be...
	About every 100 sectors	Once in 1000 times		Practically always	Minor
	1.E-02	1.E-03		1.E+00	
			UOS frequency:		Mean Accident frequency:
			1.E-05		1.E-05
6	**Result**				
	6.1 Resulting risk class	**Secure**			
	Comments on actions:				

Fill in the grey areas!

Stop
Improve
Secure
Monitor
Accept

Appendix 3: Maritime Safety Factors

Fundamental Safety Factors

- Maneuverability
- Availability of propulsion
- Controllability of ship stability
- Capability to stop ship and seakeeping ability
- Awareness of ship position in relation to the correct safe route
- Capability to maintain survivable conditions aboard ship
- Structural integrity and damage stability
- Capability to evacuate (escape routes, equipment, emergency communications)

Competencies (for various crew categories)

- Leadership and teamwork
- Communication
- Knowledge
- Application of procedures and knowledge
- Management of ship's route and related automation/equipment
- Manual steering of ship
- Ship maneuvering in port
- Situation awareness (including anticipation)
- Problem-solving and decision-making
- Workload management

Fitness for work

- Vigilance level
- Psycho-physical performance level

Procedures Practices and Culture

- Adapted to real operational situations
- Quality and clarity
- Operational planning
- Anticipating demanding operations and situations
- Managing a multitude of cultures (and languages)
- Adequate focus on safety in the presence of commercial pressures

Ergonomics and redundancy

- Usability of bridge automation (ergonomics, HCI)
- Ergonomics in how information is presented
- Adequate redundancy within the crew (deck officers)

Availability of timely and reliable information

- Aboard ship
- Between the ship and the external world

Knowing and respecting Operational Limitations

- Shipload planning and loading: stowage, appreciation of cargo characteristics, volume.
- Limitations concerning the route, speeds, etc.

External Safety Factors

- Manageability of external threats (e.g. restricted waters, fairways, infrastructure)
- Manageability of threats related to conditions (e.g. weather, visibility, ice, currents)
- Manageability of threats caused by other vessels
- Manageability of exceptional phenomena and situations (icebergs, pirates)
- Pilotage
- Icebreaker assistance
- Towage
- VTS operations
- Port operations

Appendix 4: Table 1, reviewing properties of existing risk management frameworks

	REQUIREMENTS	ICAO	IMO	RSSB	CASA	ARMS	Aven 2014	Leveson 2015	Nisula 2017
		Safety Management Manual (SMM)	*Formal Safety Assessment (FSA)*	*Safety Risk Model (SRM)*	*Regulatory Safety Management Program*	*Event Risk Classification and Safety Issue Risk Assessment*	*Methods from "Risk, surprises and black swans"*	*Risk management through leading safety indicators*	*NRMF*
RI-1	(RI-1) As the transport system is constantly evolving, and so are the risks within, the RMF needs to include provisions to feed in a continual flow of data/information from the system.	Y	N/A	Y	Y	Y	No explicit entry points for operational data.	Yes, but only related to the indicators.	Y
RI-2	(RI-2) There is already an established flow of *safety information* within the transport system. The RMF needs to be able to digest *the types of safety information* used currently and transform such data into useful risk information.	Yes in principle. No method to transform into risk information.	N	Yes, at least the typical incident data.	Y	Yes for events, and other flows could be adapted.	No explicit entry points for operational data.	N	Y
RI-3	(RI-3) The need to carry out a risk assessment related to a safety issue or a change must be one of the normal inputs to the process.	Yes in principle.	N/A	Not part of the basic process. Could be an ad hoc study.	Safety Cases can processed as an input, but the comparison is not visual.	N/A. No risk picture.	Y	N/A	Y
RA-1	(RA-1) The new risk perspective (ind. focus on uncertainty and surprises) should be embraced.	N	N	N	N	N	Y	N	Y
RA-2	(RA-2) There needs to be a specific focus on building knowledge prior to risk evaluations and decision-making.	N	Not explicitly. Importance of data is stressed, ref. "confidence" in results.	No. Based on fixed model, which is assumed valid.	Yes, using several data sources and various meetings.	N	Y	N/A	Y
RA-3	(RA-3) Knowledge of the people within the organization (e.g. the safety agency) needs to be captured, including tacit knowledge.	Not explicitly covered.	N/A	No. Limited to building the model.	Y	Not explicitly covered.	Y	Yes, during the definition phase.	Y
RA-4	(RA-4) People from outside the core group (in the agency) and operational people from the field should be pulled into the process (to capture knowledge and to create diversity).	N	N	No. Based on a fixed model. Data from operations.	Yes. Involvement of stakeholders.	N	Yes. "Experts outside the core group"	N	Y
RA-5	(RA-5) One should aim at understanding the resilience of the organizations carrying the risks, and to find ways to describe it, preferably in a comparable way.	N	N	N	N	N	Not explicitly. But importance of resilience is recognized.	N	Y
RA-6	(RA-6) The RMF should feature a holistic risk picture including risks/threats from the 4 modes of transport and allow comparisons of risks relative to each other.	N	No. FN-diagram could be used for this, but this is not part of FSA.	Partly. Single picture for railway risks.	N	N	Yes. The examples suggest how this could be done.	N	Y
RE-1	(RE-1) The decision process should be such, that the decision on the acceptability/priority of a risk or a solution can be done in a holistic manner, taking into account the assumptions and the strength of knowledge, while paying attention both on the risk and on the costs of its treatment and considering the various priorities, not only safety.	N	Partly. Only cost-benefit assessments are considered.	N	N	N	Y	Partly. Assumptions are explicit.	Y
RE-2	(RE-2) Consequently, ideally the risk picture should be able to host non-safety risks (e.g. financial, reputational risks) in a compatible manner.	N	N	N	N	N	Y	N	Y
RE-3	(RE-3) Decisions on risk acceptability need to be done by humans, due to their capability to consider ethical principles, various conflicting priorities and make value judgments; i.e. the risk management methodology does not need to (and should not) try to produce the final answers directly.	No. Answer seems to come directly from method.	No. Emphasis on data.	Yes. Method gives risk levels only.	Not specified. Decisions seem to be taken in meetings.	Partly. SIRA gives acceptability directly, ERC is more flexible.	Y	No. Actions are pre-defined conditional to indicator values.	Y
RE-4	(RE-4) The black swans need to be addressed and the high-impact–low-probability scenarios must not be dismissed solely due to their low probability.	N	N	No. Low p scenarios add up to the total risk value, but are not treated as potential black swans.	N	N	Y	Not directly, but a black swan could show that an assumption was wrong.	Y
RE-5	(RE-5) Ideally, the risk aversion policy should be left flexible (i.e. not pre-determined).	N	Y	Y	N/A	Yes. Depends on customization.	Yes, implicitly.	N/A	Y
RE-6	(RE-6) It should be possible to apply specific industry references (e.g. such as by IMO, RSSB or EASA) related to risk acceptance or severity assessment.	N	Y	Y	Yes, implicitly.	Yes., implicitly.	Not explicitly covered.	N/A	Y
RT-1	(RT-1) Risk treatment options should also be presentable in the risk picture with some measure of their associated "costs" (in a large sense), so that alternatives can be compared.	N	Partly, in the form of cost-benefit assessments.	N	N	N	Not explicitly covered.	N	Y
RT-2	(RT-2) There needs to be a way to present and properly address black-swan-risks and other very low probability risks, e.g. by building suitable resilience in the system.	N	N	N	N	N	Y	N	Y
RT-3	(RT-3) The interventions with the system need to be adapted to the nature of the system (or sub-system), e.g. referring to Cynefin.	N	N/A	N/A	N	N/A	Not explicitly covered.	N	Y
RT-4	(RT-4) As the transport system is seen as a system of sociotechnical systems, it is important that interventions embrace the typical features of CAS, including non-linearity, unpredictability, counter-intuitiveness, unintended consequences of interventions and the fact that emergent system properties like safety cannot be controlled directly.	N	N/A	N	No. But stakeholders consulted widely.	N/A	Not explicitly covered, but adaptive approaches introduced.	No. But a large part of the larger system is considered.	Y
RT-5	(RT-5) An adaptive approach needs to be adopted for risk management and interventions. A recommended modus operandi is to use parallel experiments (e.g. small-scale implementations) and to adapt them based on the feedback e.g. by expanding the successful interventions.	N	N/A	N	N	N/A	Yes, adapted approach introduced in principle.	N	Y
RT-6	(RT-6) A holistic view of the whole must take precedence over fragmented approaches because in a CAS, different risks (and interventions) are typically interconnected (Ackoff's "mess").	N	Partly. Only risk control interdependencies are mentioned.	N/A	N	N/A	Not explicitly covered.	N	Y
RT-7	(RT-7) Stakeholders need to be involved systematically to give their inputs and feedback.	N	Not explicitly covered.	N/A	Y	N/A	Yes, indirectly, e.g. discursive strategy.	N	Y
RT-8	(RT-8) While a complex system can never be fully understood/described, constant learning about the transport system, its risks and its reactions to different types of interventions should be facilitated. Such coevolution between the decision makers and the system can be seen as one of the key objectives of the process and improve future interventions.	N	N	N/A	Not explicitly covered.	N/A	Not explicitly covered.	Partly, through the indicators and assumptions.	Y
RT-9	(RT-9) Irreplaceable human characteristics (e.g. sensitivity to history, context, objectives, values) should be exploited in making sense of the CAS, but	N	No. Strive for "objectivity".	N	N	N/A	N/A	Yes, biases explicitly addressed.	Y
RT-10	(RT-10) Both from the intervention and learning points of view, three distinct levels need to be considered: events (system behavior), organizations (elements) and the system level (parts of, or the whole transport system); and besides analysis, there should be an effort to apply synthesis, i.e. study the role the system plays within the larger context, and the objectives/constraints from above.	N	N/A	N/A	Yes, at least these levels are explicitly recognized.	N/A	N/A	Partly. The system level may be incomplete.	Y
C-1	(C-1) Safety and risk management functions need to be carried out within acceptable levels of efficiency, resource and cost. This is particularly critical for process steps which are performed frequently (e.g. event risk assessment).	Y	N/A	Partly, in its limited scope of the model. But updates are rare.	Yes, implicitly.	Yes. Workload and speed are key concerns.	Yes, resource limitation is recognized.	No. Difficult to estimate resource requirements, but probably very high.	Y
C-2	(C-2) Thanks to the process, the Agency should be able to focus its resources as precisely as possible on the most critical risks (which could be combinations of risk scenarios and organizations, for example).	N/A	N/A	Yes, provided results are correct.	Yes. At least should be possible.	Yes, provided results are correct.	Not explicitly covered.	No. The approach does not deliver priorities.	Y
C-3	(C-3) In addition to inputs from Risk Identification, it should be possible to integrate into the process proposals coming from outside, e.g. from the political system, and make sure such items are comparable with other elements.	N/A	N/A	N	No, no global picture where comparisons could be made.	N	Y	N	Y

Appendix 5: Trafi safety data sources at the beginning of the project

Road transport

In terms of governance and number of involved parties, road transport was the most fragmented mode. It is also reflected in the safety data.

The main data sources in 2013 were:

- Data created by police on road accidents. The data cover all basic information related to a road accident, such as place, time, type of accident and number of victims (fatalities and injuries). The main limitation of this source is that the police is not always present at the accident scene, especially when light traffic is concerned (e.g. bicycle accidents). The location data is approximate. The data is stored in the information system of the police and Trafi did not have direct access to the system.
- The rescue operations - who are often among the first ones at an accident scene -collect data focusing on the injuries of the victims but also on the accident type and the vehicles involved. The data is available near-real-time in the *pronto*-system. Importantly, there is an initial split between severe and minor injuries – however, this assessment in *pronto* is not updated based on later more precise information (e.g. from the hospital). The location data is precise (based on GPS coordinates). Trafi has direct access to *pronto* and it is also possible to make statistical queries in the system. There is a similar limitation as with the police data: the source does not cover all road accidents.
- The Finnish Transport Agency (Liikennevirasto) – the agency responsible for the traffic infrastructure – publishes a refined version of the official accident statistics and does this through a monthly edition where data on month N would be available in in the edition of month N+2. The data is published in a useful excel-format and the location data has been checked and corrected. There are about 100 parameters recorded of each accident in this database.
- An infrastructure contractor Destia created a software called iLIITU where the above three sources are combined and the results are presented in a geographical format on a map surface. The software features various analysis functions on the aggregated data.
- Statistics Finland (tilastokeskus) publishes the official yearly statistics on road accidents. The main source is police but other sources are used to cross-check and complete the data. Accidents which were in fact due to a suicide cannot be separated from the data. This ”pollutes” the data from the perspective of traditional road safety analysis.
- The Traffic Safety Committee of Insurance Companies (VALT) publishes a yearly report which is based on the results of the investigations of road accident investigation teams. The investigations cover all fatal road accidents and present both analysis and statistics. Trafi has access to this data in an electronic format.
- The VALT report also reflects another source: traffic accident data collected by insurance companies related to compensations paid from motor liability insurance.
- Some high-profile road accidents are investigated by the Safety Investigation Authority (OTKES). The resulting investigation reports provide in-depth information on these accidents as opposed to a continuous data flow.

In addition to the above data sources, there are some other which were not used systematically in the core data analysis activity, for example the unannounced roadside inspections of commercial vehicles inspections, carried out by Trafi based on the EU law; and the police traffic oversight data, which is not available at Trafi. Finally, a lot of research on road safety exists both nationally and internationally. For example, age as gender as parameters to accident proneness are well researched.

An important recognized data need was data on the *severity of injuries* caused by accidents. The difference between severe and minor injuries is significant. The challenge was to make hospitals carry

out a suitable severity classification and link it with the correct accident. The many hurdles included barriers related to cost and privacy legislation. Another recognized gap was to get data concerning accidents not covered by police' or rescue operations databases.

The main particularity of road traffic safety data is that there are a lot accidents and even a relatively important number of fatal accidents. Because of this, there is an opportunity to guide the safety work based on real data on real accidents – an approach which in not feasible in other modes due to their very low number of significant accidents. In road safety, the main focus in analysis and safety work tends to be on fatal accidents. Consequently, there is less need to work on near-misses, incidents, etc. and try to assess the risks of similar events causing accidents in the future. In fact, the definitions of risk in road safety work are usually either:

- Number of (fatal) accidents per exposure (quantity of traffic movements).
- Number of (fatal) accidents per time period and length of road section considered.

Interestingly, these definitions do not consider the *severity* of the accident. As long as there is no common measure for the severity of minor/major injuries and fatalities, these outcomes will appear separately in different columns/bars and accidents with multiple victims cannot be compared in terms of total impact.

The day-to-day work of the road safety team in the analysis department consisted mainly of creating **status analyses** based on indicators (typically reflecting quantities of various types of accidents or fatalities) and compiling **specific analyses as requested by Trafi management or the ministry**.

Railway transport

In 2013, the only continuous raw data source for Trafi on railway operations was a flow of text messages on operational disturbances. This was not a *safety* data channel. Trafi also received a monthly safety summary from the railway operator VR, which was the only operator in Finland for commercial passenger traffic.

VR also provided some internal safety investigation reports but only based on Trafi requests. Occasionally such reports were also received from other railway contractors. The Finnish Transport Agency published about 30 safety investigation reports a year and these were available to Trafi.

The situation improved significantly in 2014 when VR gradually made available not only the summaries but also the raw data in an electronic format. This data consists of safety reports typically describing incidents or near-misses in a standard format. They cover commercial operations, shunting, track maintenance and level crossings. At the time, there were about 4000 such reports a year.

Like in road transport, a lot of research is available focusing on different railway safety related topics.

The analysis work concentrated on **maintaining safety indicators both for Trafi and for EU level statistics**. Additionally, a yearly safety report was made and many ad hoc analyses, which sometimes contained qualitative risk assessments typically on planned changes. The data sources did not provide a base for carrying out continuous quantitative risk assessment.

Marine transport

In 2013, Trafi had access to the following maritime safety data:

- Results from port controls. Trafi gets the results of Flag State Controls performed on Finnish ships and collects the observed findings in an excel table. Similarly, Finnish vessels are subject to Port State Controls worldwide and findings from such controls are collected by a body called Paris MoU and are available for Trafi. Trafi also gets the results of Port State Controls performed in Finland for foreign vessels. All the mentioned Port Controls focus mainly on technical items and on the equipment available on the vessel. They provide

information on the technical condition of the vessel and compliance with regulations concerning safety equipment and the crew. Paris MoU also creates a rough risk classification for individual ships and for ship-owners.

- Accident and incident data.
    - When a vessel encounters an accident, it is obliged to make an official report about it to its flag state. Trafi gets this information directly from Finnish vessels.
    - The European Maritime Safety Agency EMSA maintains the EMCIP database where accident/incident data is classified based on a large taxonomy.
    - Lloyd's List Intelligence is the best source for accidents within commercial maritime operations and it contains also more minor occurrences which the insurer has become aware of. Lloyd's uses indicators to profile vessels/vessel types/ship-owners/etc. in terms of safety and security.
    - Major accidents are covered by Accident Investigation Reports prepared by the investigation boards. Accidents in the Finland are investigated by the Safety Investigation Authority (OTKES).
- Vessel Traffic Service (VTS) acts as a traffic coordination body in coastal areas and reports incidents which take place within its operational zone. The data contains location plots of the vessels and a short narrative of the incident. There is no visibility to what takes place on the vessels. VTS covers the Finnish Baltic sea coast and the vessel traffic on the lake Saimaa. There are about 200 incident reports per year.
- The Finnish organization for pilots, Finnpilot runs a voluntary reporting scheme which is run on portable tablet computers. These incident reports cover the piloting phase of the operation only but are rich in detail. They attempt to identify causal factors too.

As can be seen, the most obvious data source – the vessels themselves – is missing. Different ship-owners run distinct types of safety programs and collect safety data accordingly. There were attempts to start collecting ship-owners' safety data into Trafi, but in 2013-2014 this data source was still not available.

The available sources provide interesting information from specific perspectives, but the coverage compared to the whole operation is very limited. VTS and pilot data both apply to coastal operation and there is no data source covering normal operations in the open sea. None of the sources is able to describe the operational dynamics on the bridge during normal operation (without an external pilot). The correlation between port control findings and operational safety is questionable and could be compared to inspecting a tennis player's shoes before his match and trying to predict whether he wins based on the findings.

The analysis activity focused on **maintaining the local safety indicators and carrying out various ad hoc studies and contributing to the yearly safety report**. There was not continuous quantitative risk assessment activity. Lloyd's List Intelligence does contain risk profiles and ParisMoU makes risk classifications on ships and ship-owners but this is far from a local risk assessment focusing on what would be the key maritime safety priorities for Finland. The 3-level safety indicator system had not been fully completed for the maritime domain, so the third level (causal factors) indicator definitions were still in progress.

<u>Aviation</u>

In 2013, Trafi had access to the following aviation safety data:

- Safety occurrence reports. This is by far the richest safety data flow into Trafi with about 5000 reports a year. Roughly 80% come from major airlines and the airport operator Finavia. Occurrence reporting also covers general aviation, but reports received from general aviation are typically very short.
- Trafi auditors carry out audits at operators. This creates detailed audit reports and compliance lists. There are about 15 audits at Air Operator Certificate holders every year. Current approach to audits is to observe how well the operator is able to take care of its own safety,

rather than focus excessively on operational details. A key result of the audits in addition to the official documents is the knowledge and experience gained by the auditors. Over the years they get to know the operators very well and also sends aspects like general atmosphere in the company which are not recorded in the audit reports. Due to their background and own information networks the auditors keep getting information on the operators also between audits. Audits include a discussion with the accountable manager of the organization.

- SAFA and SANA assessments (Safety Assessment of Foreign Aircraft ja Safety Assessment of National Aircraft) are assessments carried out at airports according to a standard 54-point international checklist. These checks cover items such as pilots' licenses, required manuals in the aircraft, safety equipment in the cabin in the cockpit and the technical condition of the aircraft. The checks in Finland are carried out by Trafi inspectors and the results are usable by Trafi and also sent to a European database.
- Like in other modes of transport, the Safety Investigation Authority (OTKES) investigates all major accidents and incidents and publishes the official final reports. Similarly, information on accidents abroad can be obtained through the various national investigation authorities.
- ECCAIRS is a European database for aviation safety events. According to the ECCAIRS website: "The mission of ECCAIRS is to assist National and European transport entities in collecting, sharing and analyzing their safety information in order to improve public transport safety" (Joint Research Centre 2014). In theory, this database can be used for various analyses using the different classifications available in the database. However, the data is so de-identified that it's practical value is low.
- Trafi does not have direct access to flight data collected and analyzed by operators in Finland. However, there is a national group which meets twice a year to discuss Flight Data Analysis (FDA) related issues and Trafi is an invited member in the group. This gives access to summaries and an opportunity to discuss the findings together with the operators. Some key results are also used in Trafi's indicators. FDA coverage of Finnish aircraft fleets is above 80%.
- To get visibility in maintenance activities and continued airworthiness activities, oversight data was used through the QPulse tool.
- In terms of exposure rates, Finavia is able to provide Trafi with detailed data on the commercial flights. For general aviation the information is at the level of yearly number of flights.

Due to the high number of safety occurrence reports and the requirement to classify them according to a European scheme processing the occurrences had many stages and introduced a high workload. At times, the team was unable to handle the quantity of reports at the pace they arrived, so occurrences were piling up. Occurrence reports where in different places and in different formats due to evolutions in the tools used. The ECCAIRS software in its 2013 version was rigid and primitive in terms of its functions.

In addition to processing individual reports the analysts carry out analyses using the whole available database of occurrences. The focus is then at the level of issues and themes. The tool used – Qpulse - provided fairly primitive analysis functions based on standard reports and performance rates. To manage the practical work, three excel files were used:

- A follow-up table was used to manage ongoing actions often involving people outside the department. This table included columns for issue severity and estimated repeatability.
- Table for high priority issues and events. The table was used to manage internal actions within the analysis department. Again, the table included columns for severity and repeatability.
- Data for indicators. The main system for monitoring aviation safety for Trafi at the time was the indicator system. There was a file where all events which would contribute to one or more indicators were logged.

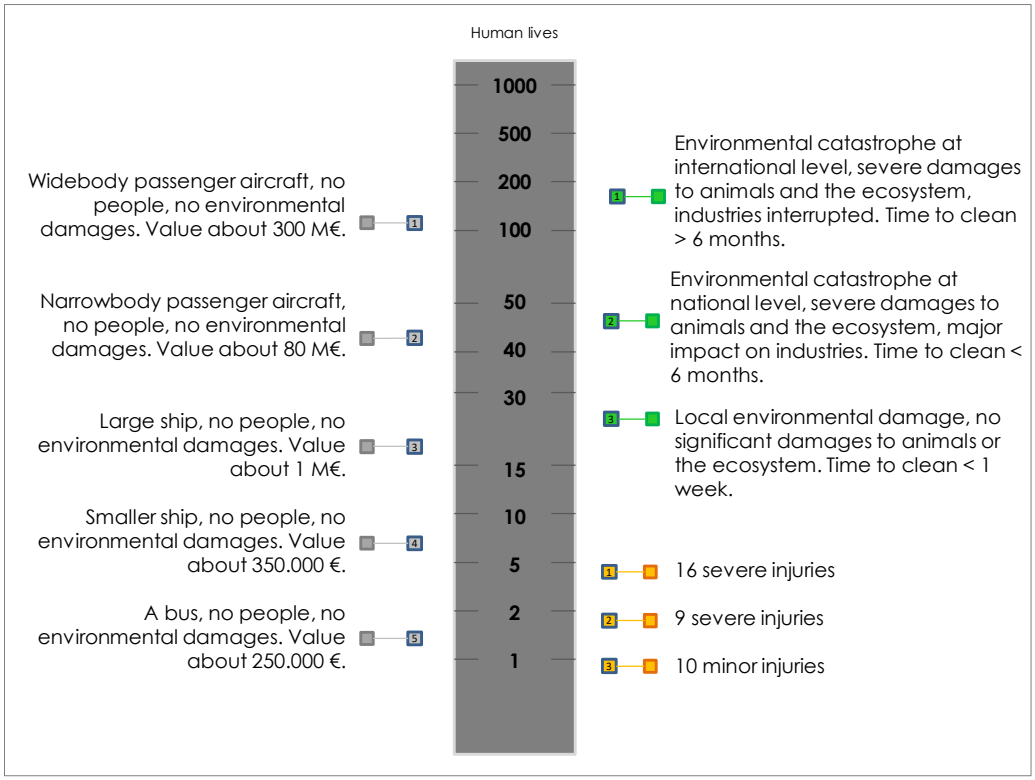
All events related to Air Traffic Management (ATM) had to be processed with the dedicated ATM Risk Assessment Tool (RAT). This tool was characterized as clumsy and difficult to use - therefore it was used only when it was mandatory based on practices in the EU.

Based on the various processes used, events got classified in different ways: severity class, occurrence category, event type, links to used indicators, responsible body, department or person. There was no official method for risk assessment nor a model that would support such a method. As its key deliverable **the occurrence analysis process led to results in the form of the safety indicators**. The indicators count how many times different types of occurrences have taken place without trying to assess the risks involved.

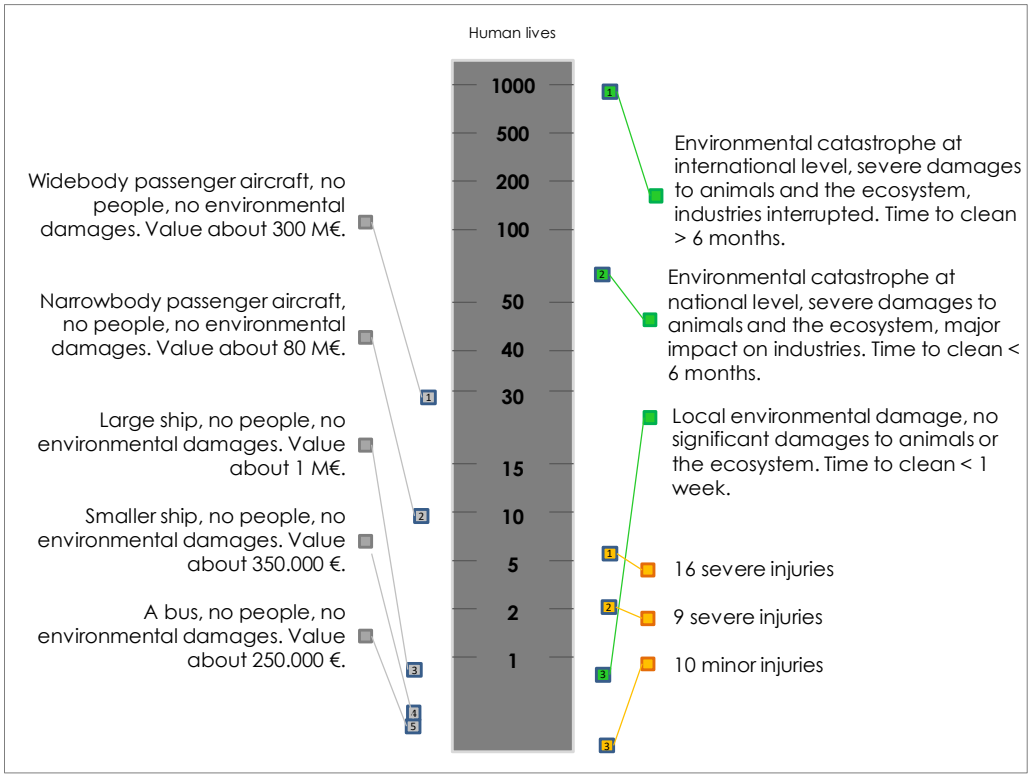
Different types of analyses were also performed and the topics for the internal review meetings were chosen and prepared. Periodical safety statistics and publications were produced and maintained.

Appendix 6: Exercise for relating different types of outcomes on the same severity scale

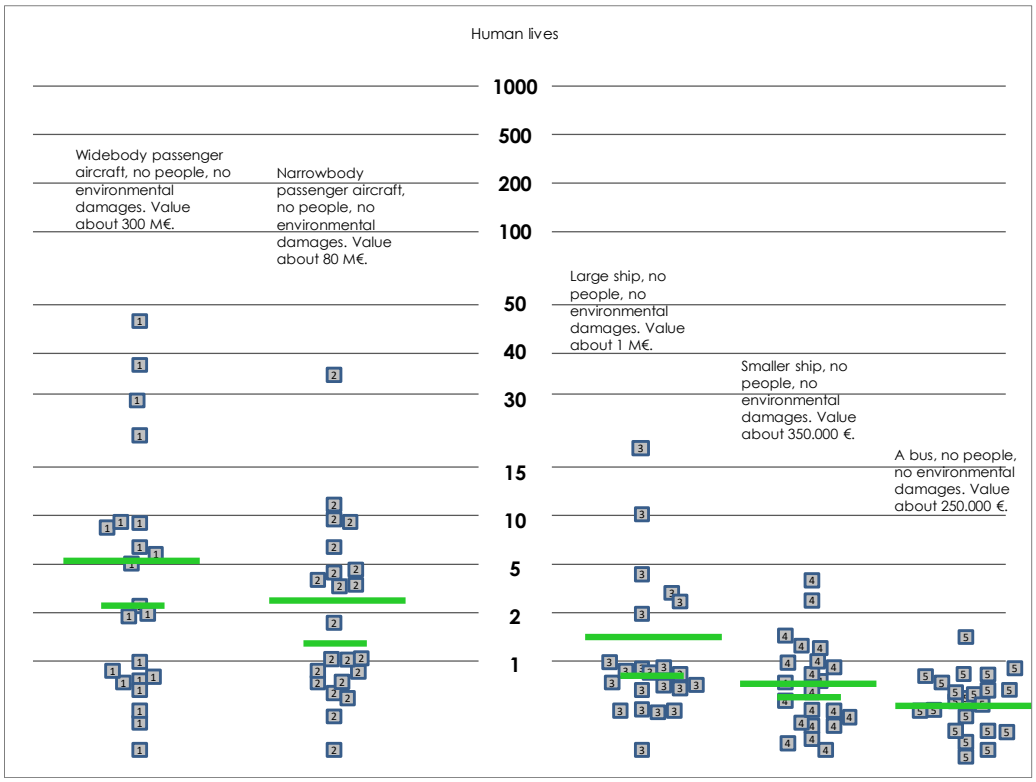
The inputs from the respondents were collected using the following PowerPoint layout:



An example answer from one respondent is provided below.



The results of the exercise are summarized on the following two slides. The mean values (longer line) and some median values (shorter line) are also indicated.



The answers on material damages and human injuries produce relatively tight clusters. This is not the case for environmental damage. The scatter in the answers seems to increase when the amplitude of the damage increases. This is the case for all three types of outcomes.