SPRINGER BRIEFS IN
APPLIED SCIENCES AND TECHNOLOGY

René Amalberti

Navigating Safety

Necessary Compromises and Trade-Offs – Theory and Practice

ICSI
Institut pour une culture
de sécurité industrielle

Springer

SpringerBriefs in Applied Sciences and Technology

For further volumes:
http://www.springer.com/series/8884

René Amalberti

Navigating Safety

Necessary Compromises and Trade-Offs - Theory and Practice

ICSI
Institut pour une culture
de sécurité industrielle

Springer

René Amalberti
Haute Autorité de Santé
Saint Denis-La Plaine
France

ISSN 2191-530X ISSN 2191-5318 (electronic)
ISBN 978-94-007-6548-1 ISBN 978-94-007-6549-8 (eBook)
DOI 10.1007/978-94-007-6549-8
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013933588

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Fifteen Years Have Passed Since the Publication of "La conduite des systèmes à risques" [1, 2]

The safety of complex systems has lost none of its currency; indeed the opposite is true. To cite a few examples, we could mention the twenty or so aviation disasters that still happen in the world every year, the chemical disasters which are almost as frequent but have longer lasting impacts, the explosion at the Total AZF plant at Toulouse in 2001, the sinking of the tanker Prestige causing an unprecedented oil slick on the French and Spanish coasts in 2002, the explosion at the BP oil refinery in Texas in 2005, the explosion at the Buncefield oil terminal in 2005, the drilling error by the Indonesian oil company which caused a mud volcano that has been flowing uninterruptedly in Sidoarjo since 2006, the explosion of the BP platform in the Gulf of Mexico in 2009 causing an oil slick that covered the whole South Coast of the USA and the rare but catastrophic nuclear disasters (Chernobyl in 1986; Fukushima in 2011), not forgetting the problems affecting public services: the thousands of deaths every day caused by an unsafe medical system or the thousands of dangerous trades made by bankers that were revealed by the international subprime crisis in 2008 and the renewed European debt crisis in 2011 (involving the disappearance of perhaps USD 25 trillion). The list is simply too long to attempt an exhaustive summary. Even more importantly than the deaths—numbers of which are tending to fall in proportionate terms—one is struck by the diversity of contexts involved and the increasing seriousness of the disasters and the huge economic repercussions that they entail.

This reveals all the foundations of a system that is maintaining a fragile balance on a global scale: producing more, using more complex tools, in more difficult places, inevitably resulting in greater risks; asking science to control this growing risk by seeking a magical alchemy to optimise control over the multiple trade-offs between contradictory risks: access to innovation and the emergence of new risks, competitive markets, free enterprise and the limitation of legal constraints, safeguarding property and human safety, immediate safety and long-term safety, the problem of waste etc.

I have spent my life studying these risks and seeking out this mysterious alchemy that would make it possible to bail out a boat that seems determined to take on more water. Above all, safety is a paradox: people demand safety once they have already taken risks.

Along the way I have often found myself pondering the problem of “how to improve safety and risk management”. This book is last in a trilogy that reflects this personal journey, which has been made up of three complementary aspects:

- understanding and improving the individual management of risks in the workplace in high-risk industries (management of high-risk systems [3]);
- changing perspectives and improving systematic risk management in business (the series of books co-edited with the MSH-CNRS in Grenoble on the seminars entitled Risks, errors and breakdowns, 2001, 2002, 2003 and a series of articles [4–8]);
- and finally, to help to control high-risk systems by providing an integrated model for the management of trade-offs in the area of safety (this book).

These three complementary aspects have emerged naturally from three successive but quite separate periods in my professional life.

- The first was a period of academic research. Although I was trained as a doctor, this period of academic life really began with my second course in cognitive psychology and my posting in 1982 to a permanent job as a researcher in a military laboratory (the Institut de Médecine Aérospatiale). I immediately came face to face with aviation accidents and focused my attention on human error. This approach fed into my theory of individual risk management and established the basis for cognitive compromise.
- The second was a period of interdisciplinary activity. In the late 1980s I had the opportunity to work closely with Airbus, Air France and the ICAO (International Civil Aviation Organisation) on the development and global dissemination of the first CRM (Crew-Resources Management) courses. Due to the resulting knowledge of the aviation sector I was seconded to the European JAA (Joint Aviation Authorities), where I worked as head of safety and human factors until 1999. In this position, I learned that safety is essentially interdisciplinary in nature but that it is understood in different and sometimes contradictory ways by each discipline. Above all I was able to see for myself and confirm that the idea that it was possible to reducing all these differences was an illusion. This is because they are based on deeply rooted models, each of which has its own reasoned approach which is in conflict with the other approaches (economic, political, human, technical and even cultural, with differences between the United States and Europe in terms of attitudes to regulation). I drew from this a way of conceiving regulations and approaches to safety and I was able to test these and apply them together with my American partners when building the common platform for regulation of human factors in civil aviation.
- The third period was one of action in the area of governance. Since the late 1990s my multiple advisory roles (on environmental risks, energy and safety in health care) and the experience of directing research programmes in the area of safety (energy and transport) allowed me to understand the view of safety

as a severe crisis surrounded by a lack of theoretical understanding, like an object blazing in the midst of an empty ocean. Business leaders, under pressure from the media and maintain a focus on the short term, are often too optimistic about their results, convinced that simply pursuing a policy of tighter controls and stiffer penalties for front-line operators will provide the ultimate solution to their problems. Meanwhile, evidence continues to accumulate that it is precisely this policy that is generating the crises feared by those same politicians and business leaders. This work is about that paradox, those consequences and the models that are available to navigate the risks.

Evolution or Revolution? Compromise and Adequacy

For a number of geopolitical reasons, the world is changing quite rapidly as we enter the twenty-first century; the new era that is beginning strongly favours those with ideas about radical change in the area of safety at both individual and organisational levels. In the area of safety, the time of traditional human factors based on human limitations and human performance is coming to an end. It is giving way to models containing dynamic linkages (Joint-Cognitive Systems) and systemic models. We are only now beginning to perceive the effects of this radical change in the area of risk management.

Tomorrow's accident, which will be rare but no doubt even more disastrous, will be an accident where the regulations were in place to prevent the problem, or perhaps where no-one actually made an identifiable error and no system truly broke down but all the components had been weakened by erosion; the degree of variation within the operating conditions will one day prove enough to exceed the tolerable linkage thresholds. Paradoxically, the safety of the system will have staked everything on reassuring procedures. With support from these it will have progressed, gained confidence, and then become weakened over time by eroded defences, increased tolerances and loss of the expertise required to manage difficult situations. The new idea of resilience must be understood in these terms: the increase in controlled safety which is imposed by regulations necessarily takes place at the cost of increased rigidity, a desire for tremendous standardisation of both technologies and human beings, ultimately resulting in operators who are less capable of adapting to surprises (this has a negative impact on managed safety, which is based on the expertise of operators and can be linked to the idea of resilience). The art of successful intervention in safety involves controlling the compromise and the trade-offs between the benefits accruing from controlled safety and the resulting losses in terms of managed safety.

The idea of compromise underlies the entire theoretical and practical structure set out in this book. Compromise is two things. It is the necessary cognitive or intellectual "micro-centered" compromise that the worker must constantly regulate, between external demands, his own know how, competing tasks and motivations and his own physiological state of fatigue and stress. It is also the "macro-centred" compromise of control, which models more or less conscious trade-offs between performance and safety at the level of the management of complex systems.

- The first compromise, which is described as micro-centered, takes place at the level of the operator. It brings us back to one of the last and hugely misunderstood points of human psychology, since it mobilises the system of intellectual control and is of course variable and subject to revision, making it largely inaccessible to the methods of study used in traditional psychology, which postulate a degree of stability in order to capture and measure a baseline in terms of intellectual ability (memory, attention, vigilance etc.). Having said this, considerable progress has been made in psychology, and although it is not possible to characterise the control of cognitive compromise on a moment-by-moment basis, it is now possible to characterise the variables that modulate it almost in real time. All the work that has been done in the past on the operation of high-risk systems was focused on this modelling process and the very significant consequences in terms of designing safe operator assistance systems. These are included in this book in the form of a summary.
- The second compromise, which is referred to as macro-centric, forms the innovative heart of this book. It concerns the trade-offs that take place between performance and safety when navigating risks at the level of company management. Sometimes these are described in terms of deciding to make specific sacrifices. Outrage over compromises of this type represents the daily grist of the tabloid press. This book offers the keys to those trade-offs and the sacrifices that are required. Almost all the keys to this compromise are surprising and at first sight they may at times seem politically incorrect, but in the end they are easily understood on further reflection. Some of the most paradoxical results that the reader will discover include the idea that the institutionalisation of safety is an emergent property of systems that are already safe, and that strategies for intervention should make use of methods and tools that differ significantly depending on how safe the enterprise is, while these should impose increasing burdens as safety improves. In fact the level of safety has the surprising property that it is never adequate and it actually generates societal demand which increases in parallel with the progress that is made. It is a variable with no maximum, and all improvements result in increasingly severe judgements by outsiders on the small number of safety failures that still occur. The reader will also discover that choosing an inappropriate safety strategy, especially one that is too ambitious, ultimately causes safety failures.

There is a surprisingly high degree of continuity between the models of the two types of compromise at the micro and macro levels or at the individual and organisational level, as if the same models perhaps form part of a fractal perspective. The psychological performance of operators under real everyday conditions, the level of attention that they maintain and the relevance of their choices, always appear imperfect to the observer and almost disappointing in terms of what they believe these people are capable of doing and even in terms of what they have seen for themselves during training. The performance of enterprises in terms of safety is almost always disappointing too, and always lower than what one would expect from the debate and when considering the organisations themselves.

The judgements made in both of these cases would be hasty and imprecise. Of course what is achieved does not represent maximal performance, but it is still adequate in terms of the demands of working standards. For example, a driver is not required to drive exactly in the centre of his traffic lane (although this would appear to be ideal); on the contrary, his environment is constructed in such a way that he can drive within a wider space, even making use of a degree of additional tolerance if he does not see anyone approaching him. This structure, which is shared, conceived and created both through technical consensus (decisions on road width and road ergonomics) and by social consensus (verbal instructions given to the police on behaviours that should and should result in a reprimand) makes it possible to come up with a reasonable response to the demand for safety. This comes back to the idea of "adequacy": "The driver's social contract is to stay on his side of the road". No more is demanded of him, that is "adequate" and he will not be given a penalty as long as he achieves that. This perspective is applied throughout the world of work, and paradoxically it serves to promote rather than reducing safety, as one might expect from the laxity that it suggests: in fact the wider the road, the less the driver is required to concentrate on driving, the less fatigued he becomes and the greater the margins available to him to recover from an unexpected event or surprise (*building adequacy also builds safety*).

The suboptimal nature of this situation makes it possible to compromise, work in parallel on other aims (private thoughts, other areas of interest etc.) or even save oneself and avoid unnecessary effort, storing up energy to enhance other dimensions of performance, improve staying-power (allowing a person to be just as effective when leaving work to go on holiday as he was when returning to work in the morning), or to be able to respond to challenges and setbacks. In short, it makes it possible to live in symbiosis with a wider range of demands from society, which can never be simply summarised in a SINGLE unique objective such as immediate safety. That is because safety has never made an enterprise survive; lack of safety can kill it, but safety can never be presented as the only goal worth pursuing. This concept of ADEQUACY naturally complements the concept of COMPROMISE and it will be set out explicitly and debated in several places in this book in order to understand its definition and above all its implications when it comes to building regulations and auditing enterprises. How is it possible to successfully create this alchemy: a technical and social consensus on what is "adequate" when defining and interpreting standards?

As always, a work is never achieved by a single person, and I must begin by thanking those who have accompanied me on this long path for their ongoing contributions to this debate; in France: Jean-Michel Hoc, Jean Pariès, Maurice de Montmollin and Jacques Leplat, and internationally: Jens Rasmussen, Jim Reason and Eric Hollnagel. They have all continued to be close contacts and a source of both criticism and inspiration for my work. Next I must mention all my doctoral students, who carried out a significant proportion of the field work (with special thanks to Gaël Morel and his work on professional deep-sea fishing). I cannot mention all my other international friends from all the academic and industrial contexts in which I have spent my time and whose experience and models have made an impression on me. In particular I must mention my involvement, right from the outset, in the Institute and then the Foundation for an Industrial Safety Culture (ICSI/FOCSI) which has allowed me not only to penetrate more deeply into the secrets of major industries and the safety challenges facing them but perhaps more importantly to work alongside people with a wealth of experience; in particular I am thinking of René Deleuze and Ivan Boissières.

One of the most valuable aspects of that experience is no doubt the wide range of different influences and environments, ranging from research in the French-speaking and Anglo-saxon traditions to political governance, aviation, health care, road safety, professional fishing and major industries. In the end science is often simply about creating a synthesis. Perhaps the greatest contribution of this work is precisely that it creates a way of reflecting that brings together different approaches and contexts which normally do not interact, which caricature each other and each of which considers its own case to be so special that it cannot learn from the others, while an outsider's eye can perceive common theories and properties emerging from the whole.

Another effect of the age of this book is that its potential readership has grown and it will be available in three languages: French, English and Spanish.

In short, this book seeks to provide the keys to safety in high-risk systems in the twenty-first century, making the models as accessible as possible while maintaining an adequate degree of scientific precision, for an intended readership including teachers, consultants or industrialists who may be able to make use of the models.

The theoretical path may never reach its end, but it has at least been marked out, as has the vision encompassing the transverse links across the different trends in the world. The reader will also find a list and critical reading of a very large number of references.

The practical path to intervention in safety has also been marked out but it has been intentionally limited to general principles and a toolkit has not been included. I will leave the work of developing such tools to my colleagues. Maintaining a distance from structured assessment kits in this way is not yet another scientist’s strategy to avoid the problems on the ground; without minimising their importance, the refusal to emphasise them makes it possible to place the emphasis elsewhere instead: specifically on the political decision-making process that occurs prior to their use. Success in this area of risk management is above all about high-level strategic choices rather than measurement tools or questionnaires which are easy to use but often mask the most important aspects.

So fasten your seatbelts, dear reader: you are about to learn many things that are counterintuitive, some of them disturbing and others reassuring, but all essential to successful risk management. I hope you will enjoy reading it, and criticisms and discussion are always welcome.

René Amalberti

References

1. Amalberti R (2001) La conduite des systèmes à risques, 2nd edn. (trans into Spanish). PUF, Paris
2. Amalberti R (2009) El control de los sistemas de alto riesgo. Modus Laborandi, Madrid
3. Amalberti R (2001) La conduite des systèmes à risques, 2nd edn. PUF, Paris
4. Amalberti R, Fuchs C, Gilbert C (2001) Risques, erreurs et défaillances, Vol 1. MSH, Grenoble
5. Amalberti R, Fuchs C, Gilbert C (2002) Conditions et mécanismes de production des défaillances, accidents, et crises, Vol 2. MSH-CNRS, Grenoble
6. Amalberti R, Fuchs C, Gilbert C (2003) La mesure des défaillances et du risque, Vol 3. MSH-CNRS, Grenoble
7. Amalberti R. (2001) The paradoxes of almost totally safe transportation systems. Saf Sci 37:109–26
8. Gilbert C, Amalberti R, Laroche H, Pariès J (2007) Toward a new paradigm for error and failures. J Risk Res 10:959–75

Contents

Chapter 1
The Demand for Safety and Its Paradoxes

Abstract The world demands ever-increasing safety. That demand, however, varies from one system to another as its life-cycle progresses. The pressure on safety often reaches its maximum at the end of the cycle, which is paradoxically the time when the system more or less reaches its apogee in terms of the level of safety.

A World Demanding Safety

Problems surrounding safety in our society have never been so widely discussed. It is not so much the number of accidents but rather their immediate, high media profile throughout the world that strikes fear into the hearts of citizens in rich societies who have everything to lose in terms of the collapse of their comfort, health and values.

Worse still, there are a growing number of prophets of doom (or whistle blowers) de Kersasdoué [1], who receive intensive media attention as they broadcast forecasts of woe and confusion, seeing every accident—particularly when these have multifactorial causes—as symptoms of a society that has lost its wisdom and self-control (industrial hazards, natural disasters, forms of "suicidal" behaviour in society such as smoking, alcohol or road deaths, economic upheavals). It is hardly surprising that such a situation has turned this issue into an inescapable variable which is acquiring greater public policy—and electoral—importance.

Public bodies at the national and international level have responded to this situation by creating agencies, new authorities and offices dedicated to safety; these bodies have passed numerous laws and decrees on safety; research funding has flowed rapidly into this area (research funding in this area in Western countries has increased by almost 300 % since 1992). The same inflationary tendency is seen in University and continuing training courses, while small and medium-sized enterprises and specialist consultancies in this niche of safety intervention and consultation are becoming more numerous (+72 % in the United States between 1985 and 2000) to meet the exponential demand for audits and risk assessment reports (+430 % in 10 years).

It is this emergence of the concept of "safety" which has finally given scientific credence to a subject which was long treated simply as a variable associated with technical development.

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology,
DOI: 10.1007/978-94-007-6549-8_1, © The Author(s) 2013

Not every aspect of this public awareness is beneficial, however; the market that has been created is not only lucrative but it also has a polemical streak and still suffers from a number of vulnerabilities.

In many cases systems are beginning to operate at two speeds. On the one side there is an apparently virtuous approach to quality and safety, operating from a comfortable, cocooned environment in its new offices, bursting with certainties about the constraints that must be imposed upon the system. On the other is a production system pressurised by a market, in which the FBC rule (Faster-Better-Cheaper)[1] is the only way of ensuring commercial survival and success. The modelling of decision-making on sacrifices and the management interfaces that are needed to calculate trade-offs between safety-related and production-related values are some of the least well-known aspects of safety management. It must be recognised that the economic death of an organisation is still an accident; and the antithesis between formal safety and economic survival is perhaps not so great as might be imagined. Crucially, however, there is no scientific, global or systemic perspective on this issue.

In short, we have made rapid progress on the details (the mistakes made) and even faster progress on simple ideas and local safety measurement systems (which can be controlled and marketed) but we still have almost no model at all which is able to serve as a framework for the strategic management of safety.

Have we gone too far without taking into account the limitations imposed by the terrain? Have we progressed too fast? Is the threat of impossible levels of safety in high-risk industries as severe as has been stated? Does this represent a real threat to society? Can safety be treated as a generic problem, independently of its source? What scientific underpinning is available?

Finally, is this a scientific field in which the knowledge required is primarily technical, or is it essentially a social science?

Question of Perimeter: Which Systems are Involved and for Which Safety Process

This book deals with safety (failures) in major sociotechnical systems (energy supply, public transport and services such as banking and medicine). The safety of these systems is an issue that has lost none of its currency—indeed the opposite is true.

The list of disasters is too long to attempt to reproduce it in this work. Even more than the number of deaths (which are actually tending to reduce in proportional terms), it is the wide range of different contexts involved and the growing severity of the disasters that captures the imagination.

[1] Faster-better-cheaper (FBC), an expression from NASA which made it into a slogan before the second space shuttle accident.

These “major sociotechnical systems” all have three characteristics:

- the processes that need to be controlled are dynamic; they evolve independently but can change direction as a result of human intervention;
- they remain under the control of people who are “in touch with the process” and people within the “management loop”; the first group of people are referred to as being at the “sharp end” while the second group are at the “blunt end” [2]. This property of horizontal and vertical human control remains present at every level of technology, although the number of operators does tend to be smaller in situations involving innovative solutions;
- they are at risk; the risk is physical death for the players and/or the death of the system itself (specifically this means economic destruction). Such deaths may be isolated but they are more often accompanied by collateral effects.

There is No Shortage of “Common-Sense” Solutions to Make Complex Systems Safer

Safety is created by taking risks, just as water entering a boat creates a need to bail it out. The idea of reliability is clear in this sense: there are dangers in the world that affect the activities of human beings (energy, materials, objects and products). The risk corresponds to the frequency of exposure of human beings to these dangers and the consequences of that exposure; the risk is negligible and no doubt can be accepted for a danger which is very serious but highly improbable (injury by a meteorite) or a danger which is frequent but has only minor consequences (about of indigestion after eating too much chocolate). It is unacceptable, however, when it comes to a frequent danger with severe consequences (a carpenter cutting off a finger) or a less frequent danger with major consequences (a nuclear accident).

In those cases which are judged to be unacceptable, safety first of all involves reducing the exposure to the risk whenever this is possible and then protecting operators and citizens from that risk, whenever exposure to the risk is considered inevitable for commercial or public reasons. High-risk systems fall into the second category, where people agree to be exposed to the risk (to gain other benefits) and where the purpose of safety is to find solutions to avoid—not the risk—but an accident.

For thousands of years, hopes of continuous improvement in safety have been based on a series of fundamentals which have not varied much, although they appear to be self-evident, common sense principles. Five recurring themes can be identified in our thinking and our actions.

First idea: for the responsible body (the company, the industrial sector), the safety process should satisfy four aims, the first two of which are public and measurable while the second two are more personal:

- reduce accidents (achieve measurable progress);
- demonstrate to customer(s) and society at large that there is a willingness to do better, and that this is shown by the constant efforts being made;

- soften the critical postures and attitudes of observers, citizens, customers, neighbours in the hope of gaining recognition for the efforts made and the results achieved;
- minimise the spectre of major post-accident socioeconomic crises (for example the political crisis resulting from contaminated blood products in France, mad cow disease in the United Kingdom, the questioning of NASA after the Shuttle accidents, the huge political instability following Hurricane Katrina in the United States or the Fukushima disaster in Japan).

Second idea: the road to safety involves well-known actions that have been demonstrated in industries and sectors that have become safe (the "good students" such as the nuclear industry and aviation). We live with the idea of a system of solutions and a final common pathway for safety that is thought to be common to all industries: the highest level of safety for all can be obtained, it is thought, by putting in place the same tools that are considered to be effective in different successful enterprises.

Third idea (a refinement of the previous one): conformity with an "ideal" procedural process is a necessary part of the pat to improving safety. Reducing the diversity of professional practice through the introduction of procedures will naturally result in improved safety. The use of feedback, and the active participation of both the professionals involved and the management in monitoring deviations from procedure and the wider quality process are main tools used to create conformity.

Fourth idea (with sincere thanks to Reason): errors—which are by definition involuntary—are tolerable and in any case more or less inevitable, although the debate has not yet concluded on whether or not these are susceptible to safety interventions and a reduction in their frequency. The system should accept errors as the inseparable flip-side of human intelligence and should develop strategies to manage these errors while preventing their consequences. The barriers can be arranged into three complementary sectors: Prevention (avoiding the risk), Recovery (managing the risk and avoiding the accident), and Mitigation (accepting the accident but not dying from it, reducing the impact). The distinction between patent errors (by front-line actors) and latent errors (in the management of the system) is another vital point put forward by Reason, which underpins the importance of certain actions that target the management and not only the front-line actors.

Fifth idea (a consequence of our legally based societies?): errors may be tolerable, but voluntary departures from the norm and violations are still considered as deliberate and unjustifiable acts.

Unfortunately the path of "common sense" in relation to these five ideas is strewn with pitfalls, both in the way they are used and the benefits derived from them. The benefit of each of these ideas can be verified, but it is often limited to a relatively restricted field of application.

In order to understand this paradoxical phenomenon, this first chapter proposes a macro-scale analysis of risk management, studying the models that describe safety-related changes in industrial cycles and how they progress towards their peak level of safety. A number of important consequences are suggested in terms of macro-scale risk management. These consequences then serve as guides throughout the rest of the book.

Life-Cycles of Socio-Technical Systems and the Paradoxical Positioning of Safety in Time

In the end, all systems, whether they are biological or technical, will die. The need will continue to exist or the function will continue to be carried out, but the means by which those needs are met and those functions are carried out are subject to regular changes.

For example, mobility and trade have been keys to commercial activity and a constant need for societies since ancient times. One of the most recent means of transport to appear was air transportation. Dirigible balloons represented one method for doing this; piloted aircraft were a second method; drones or automated aircraft are already a third method. Each of these methods satisfies the same need and corresponds to a socio-technical system, a specific organisational link, and each one passes through a particular life-cycle. It is easy to show that the length of these life-cycles is now essentially the same as a human lifetime: between 30 years for the ones that die prematurely, and a little over 100 years for the most robust systems. These life-cycles used to be longer in the past and sometimes extended as long as several centuries, but the acceleration of innovation in our society has permanently speeded up the system.

In each cycle, the specific system goes through different phases from birth to ageing. The cycle model presented in this section is inspired by a number of works, particularly those of the sociologist Hughes [3] whose principal field of study over a long period was the development of the electric power industry.

The model describes three phases (Fig. 1.1) [4].

A general model describing the stages in the life cycle of major industrial systems. Major systems are also born and die, on their own scale—these

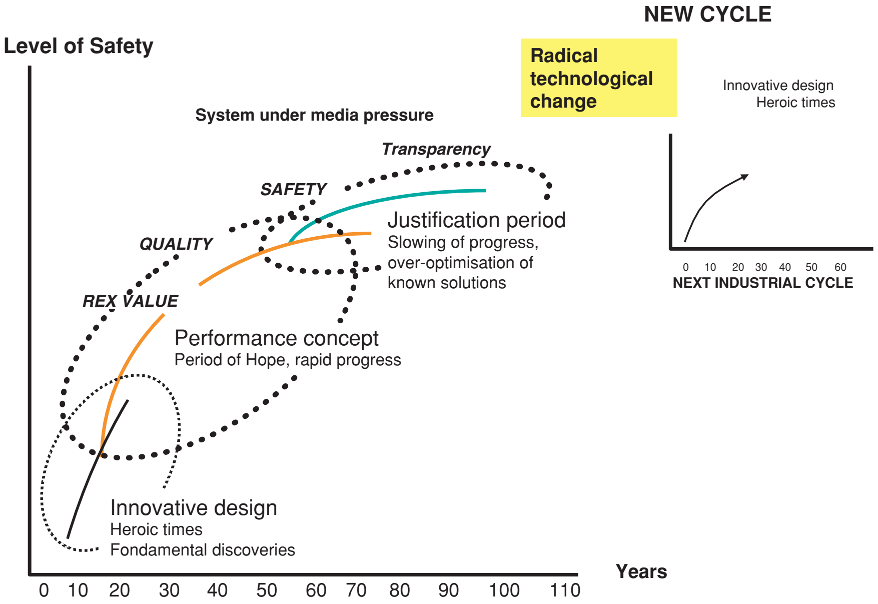


Fig. 1.1 Conception and cycling of socio-technical systems and legal vulnerability

are described as industrial cycles—and they have a life-cycle that can largely be reproduced from one system to another [3]. The death of the system does not, however, mean the death of the system's function; the function is reborn from the ashes in the form of a new technology emerging from a radical technological change: the cycle begins again, often with a different form of linkage (Fig. 1.1).

An Initial Creative Phase, Often Unseen by the General Public

When new systems are born, the innovation is carried forward by only a few individuals; safety is not a priority at this stage, even though accidents are very likely to occur in the initial phase during which people get the invention under control (immaturity). It should be said, however, that these accidents generally kill their inventors and have only limited collateral effects, while the number of prototypes is necessarily small or may even be secret.

The discovery of blood transfusion and the beginning of an industrial life cycle. Blood transfusions became possible after 1900, following the discovery of the ABO system and blood group compatibility by Karl Landsteiner (innovation stage). The period from 1910 to 1930 saw the first steps towards optimisation: progress was concentrated on the conditions for large-scale commercialisation, the organisation of the blood collection network, blood banks and the development of clinical applications.

An Optimisation and Economic Benefit Stage, by Far the Longest, Known as the "Quality Period"

This extends over several decades. This is the heart of the useful lifetime of the system, between the birth of the invention and the sclerosis of old age. At this stage, the industry accepts the innovation, standardises it, optimises it and its use becomes widespread and generates profits (in economic, ethical or welfare terms). This phase comprises two periods. During the first period, the key problems of the innovation are resolved, making it commercially usable. During this stage, safety evolves very rapidly, in parallel with each technical improvement.

The example of cycles in air transport

After the first pioneers in the late 19th century with their balloons, and the difficult emergence of a civilian version that could be developed on a large scale (1899–1906), for almost 30 years dirigible balloons, particularly those made by German industrial company Zeppelin, represented the pinnacle of technology making air freight and air passenger transport possible. This solution very quickly reached the end of its life since it failed to find an alternative technology to the use of hydrogen, which was highly inflammable. Aided by the media and political pressure, the disaster in which the Zeppelin Hindenburg was destroyed in 1936 brought this technology to a sudden end. It gave way to aircraft, which offered an alternative technology for the carriage of passengers on a large scale. For 20 years (1935–1955) aircraft on scheduled flights were objectively more dangerous than the last Zeppelins, but they nevertheless underwent global growth and an unbelievable pace of innovation (arrival of jets, automated systems etc.) that gave them credibility in terms of their potential (positive value of hope) and permitted the (numerous) accidents that still occurred during this period to be tolerated. (This was no longer the case for a technology that had been used prematurely like the Zeppelin).

Paradoxically, this cycle which began around 1935 is again coming to an end. Traditional civil aviation with pilots in the cockpit is in its final stages and can be expected to undergo a large and fundamental technological change in about 2030/2040 with the expected arrival of a fully satellite-guided system. Professions will change, as will the whole economics of the system, but people will continue to use aircraft to travel from A to B (continuity of function).

At this stage in the commercialisation of the innovation, the brakes are placed on research far upstream, the inventors are given honoured positions but are actually kept at a distance from the enterprise, giving way to engineers and commercial staff who can transform the initial idea into an effective, reliable commercial product. This prevents too much defensive action by inventors seeking to protect their “baby” and excessive chaotic activity resulting from alternative inventions that might lead to a loss of focused effort, consume resources and confuse the commercial image of the product which is being launched.

The second period corresponds to the quality period; the quality process increasingly optimises the product, introduces the concept of feedback, rigorous procedures or protocols to satisfy customers; it gradually eliminates the initial defects to optimise customer satisfaction. Regulatory constraints increase in parallel with this process. Hale et Swuste [5] identify rules at three levels which also represent three characteristics of risk management policies at this stage:

- introduction of rules governing the product, which define the targets that must be reached or must not be exceeded depending on the variable in question, for example a maximum concentration of permitted toxicity or compliance with the

ALARA (As Low As Reasonably Practicable) rule, or the statement in multiple aviation regulations that the situation should not require pilots to show exceptional skill;

- introduction of rules governing the process and requiring the creation of and respect for a control organisation (control by whom, over what, when and where). For example, airlines must have a feedback system administered by an independent body;
- rules must also be introduced to govern the process that demand specific results: for example compulsory wearing of goggles by operators or licensing and medical examinations for personnel.

Accidents become less and less numerous. Safety continues to grow in parallel with these optimisation measures but it is not governed by a specific agency or division; it simply results from continuous optimisation. The reason is simple: throughout this phase, the legal environment is still relatively non-threatening. This is a privileged period of symbiosis between economic results, public satisfaction, safe progress and tolerance of accidents. Prosecutions are rare and when they do occur, the resulting convictions and compensation payments are moderate. The defence (for the industry) is able to plead successfully that the accident occurred due to a lack of sufficient knowledge. What is more, the situation that caused the accident can no longer be reproduced at the time of the lawsuit, simply because of discoveries and progress made between the time of the accident and the time of the lawsuit (usually several years).

A Final Optimisation Phase, Known as the "Safety Period"

Progress slows and the system becomes ballistic (Hugues speaks of "uncontrolled momentum"), often relying on local over-optimisation to the detriment of an overall strategic vision. The number of regulations increases rapidly and the nature of those regulations is increasingly responsive to events. The cost of such an increase is usually a rise in the number of legal disputes and regulations that offer little benefit.

The legislative process becomes increasingly burdensome and complex Rhode [6] and legal rules multiply disproportionately in Western societies, echoing increasingly stringent public demands. The "recueil annuel de l'assemblée" (the record of new legislation in France) ran to 1,300 pages in 1990; it was 2,500 pages in 2006.

At this stage, accidents have become rare or even very rare but they are much more intolerable to public opinion. The burden of management swells; safety divisions replace quality divisions or new ones are created alongside them. Crisis cells come into being because crises themselves become more and more frequent and devastating. This situation generates two paradoxes which are difficult to manage and will ultimately cause the death of the system:

- the worst safety problems (in terms of the image of the enterprise) are perceived to exist at a time when the efforts being made to promote safety have never been so great and accidents have never been less numerous. Systems that have become very safe therefore simultaneously become more fragile;
- society demands greater transparency about the real risk, but does not know how to manage that transparency; access to this information in turn makes civil society more intolerant of all the problems that do remain; the more information is available to the public, the more it begins to doubt what is being concealed from it behind that information. This attitude of mistrust makes it difficult to develop information systems that are too transparent and in turn hampers real improvements in safe practice.

A good example of such paradoxical reactions at a late stage in the cycle

During the month of August 2005, French television broadcast images of a number of major aviation accidents over just a few weeks: the series began on 2 August with the accident involving an Air France Airbus A340 in Toronto, where there were miraculously no victims thanks to a successful evacuation on the ground; the next was an accident involving a chartered aircraft that ditched in the sea off Palermo after both engines shut down (Tuninter ATR-72), fortunately again with survivors (23 out of 39); there were suspicions of refuelling with low quality fuel, which was hardly reassuring in terms of confidence in charter companies. Barely a week later, another charter flight accident occurred on 14 August, involving a Helios Airways Boeing 737 which crashed into a mountain near Athens; apparently the pilots had suffocated due to a completely incomprehensible lack of oxygen; and only a few days later, on 16 August 2005, the last and most emotional of the accidents in this series occurred. This accident involved West Caribbean Airlines charter flight 708, carrying French passengers from Martinique, virtually all of whom came from the same village. The aircraft crashed in the Venezuelan jungle due to appalling weather conditions and very poor management by the crew. There was a huge outpouring of emotion from the general public, who had already been sensitised by the Flash Airlines tourist charter accident the previous year, which had occurred on take-off from Sharm El-Sheikh and had killed 135 French passengers returning from their holidays.

The polemical atmosphere rapidly intensified following revelations about the maintenance of charter aircraft, flight authorisations and the lack of information on dangerous airlines. The DGAC (Direction Générale de l’Aviation Civile, the French CAA) quickly announced stricter controls on charter airlines and made available a list of charter airlines “to avoid” (blacklist).

During the months that followed, passengers on charter airlines displayed extreme attitudes; on a number of occasions, having been informed of a malfunction that had caused a delay, they refused to board or asked to disembark, on each occasion resulting in disputes about the flight and ticket reimbursements. Paradoxically, technical analysis of these cases showed that the companies were in fact applying the regulations strictly (which they had not been doing previously); instead of accepting minor risks, the crews were following the procedure properly and informing the authorities and passengers in complete transparency.

Two years later, the companies in the hot seat who had made efforts, particularly the charter airlines, annoyed at being subjected to this lack of understanding, had in some cases reverted to their former approach of disclosing as little as possible to passengers and the authorities.

During this final phase, one also sees the intensified use of the tools and methods that had hitherto been shown to result in improvements (quality processes, risk elimination, various constraints on processes, protocols, regulations). The negative effects of these over-optimisations have the effect of shifting dangers. Errors of commission are replaced by errors of omission. The system uses up more and more of its resources controlling itself. This consumption, which is barely perceptible at first, becomes significant as the system develops, particularly in the form of tensions in the area of personnel management. The use of continuous quality tools, and in many cases derivatives of such tools, offers a good example of this paradox. Regulatory constraints also create orders of priority between the problems being dealt with and fail to address sectors which are objectively more dangerous but are less emotional, less subject to regulatory pressure and technical controls, and which de facto attract less media attention.

Media pressure and irrational investment late in the cycle France invested around 50 million euro in the (high media-profile) aftermath of the accident that occurred in the Mont Blanc tunnel[2] (39 victims, 24 March 1999) to improve safety in a small number of major Alpine tunnels which have claimed no more than a few victims over 20 years. This was significantly higher than the total amount invested by the French State during the same period in road safety as a whole (although more than 8,000 people

[2] Summary and analysis of the accident available on http://www.mace.manchester.ac.uk/project/research/structures/strucfire/CaseStudy/HistoricFires/InfrastructuralFires/mont.htm.

were killed on the roads each year[3]). In the early 2000s fire prevention requirements were tightened up in French care homes, following a small number of disastrous accidents that received intense media attention. The increasing budgets required to achieve compliance in this area, impinged on the shared resources of the remaining safety budgets in these institutions and these were reduced by the same amount, although issues of safety in health care, which generally do not attract media attention, were estimated to have claimed far more victims than these (rare) fires in institutions.

This growing intolerance towards the end of the cycle lowers the threshold for triggering crisis situations. Crises are more and more numerous, more unpredictable and more difficult to control. The trigger for safety crises within an enterprise is particularly sensitive to the multiplication of complaints and the phenomenon of emotional resonance generated by the media.

Finally, lawsuits are showing a noticeable shift in judges' perceptions in regard to responsibility for industries and public systems. The remaining scope for spectacular technological progress has been reduced. The general public, and the justice system which echoes society at large, tend to consider that the perceived risk is no longer linked to an innovative technology or to insufficient knowledge, but rather to an incorrect trade-off due to areas other than safety being prioritised.

In this increasingly unfavourable context, the only people who minimise the integrity of the threat and realise it too late are those working in the sector. Defence mechanisms bias their perceptions and lead to industrialists and social partners becoming locked into a double trap: corporatism and protectionism. The actors involved refuse to address the fact that the cycle is ending and invest in the next cycle; in this situation, the industrial actors who have been leaders throughout the cycle often disappear, making way for the new industrial actors who will lead the next cycle. Although this is not a true safety cycle, the example of the end of silver-based photography and of Kodak, which is at risk of bankruptcy, is almost a caricature of this locked-in behaviour.[4]

The End of the Cycle and Death

The end of the cycle is heralded by growing financial investment associated with pressure to prioritise safety policies (these may give rise to direct costs in terms of personnel or resources or indirect costs in terms of reduced activity and inability to

[3] http://www.fiafoundation.org/publications/Documents/road_safety_in_france.pdf. This distortion, which was true during the period referred to, was fortunately reduced when road safety became a major presidential issue following the re-election of President Chirac.

[4] Kodak Teeters on the Brink, Wall Street Journal, January 2012, 4, accessed at http://online.wsj.com/article/SB10001424052970203471004577140841495542810.html.

achieve safety standards). Profits fall and the system looks for escape routes, usually by restarting upstream research.

Once the conditions for an alternative are reasonably well in place, all that is needed is one last accident that attracts significant media attention (a “big one”) to cause the previous system to collapse and make room for the next system. In this sense the “big one” is not an accident which is simply exploited because of what it objectively represents in terms of its technical, organisational or human cause, but an accident that is symbolically exploited because of its systemic value.

This is how industrial cycles end. There is no shortage of examples.

The end of blood transfusion announced [7]. Blood has become very safe, so safe in fact that blood products are now tending to be in short supply and their cost has risen explosively. Upstream research was restarted as many as 10 years ago to find blood substitutes (synthetic haemoglobin) or quasi-natural palliative solutions (stem cell cultures that produce red cells). Given the potential for innovation that exists, it is possible to imagine that the cycle of using natural blood for blood transfusions is approaching its end and a new cycle is preparing to begin very soon. The life cycle of human blood transfusions has continued for just over 100 years.

At the End of the Cycle, Accidents are Often More Severe, More Intolerable and More Expensive to Remedy Legally

At the final stage in its development, the paradoxical situation has been so severely aggravated by improvements in performance that the risks associated with the very small number of remaining accidents become much higher.

The benefits offered to consumers are constantly increasing; depending on the technology, the gains are associated with reduced prices (energy, cars, agrofood), more efficient service (faster, more convenient, public transport) or innovation that creates possibilities that previously did not exist (chemicals, medications, surgical techniques).

Technological progress makes it possible to exploit areas that had previously been inaccessible or neglected: people travel faster and produce more, traffic flows increase, complex treatments are given to people who had previously been written off and excluded from the medical system. The first paradox is that this process of progress necessarily results in new risks.

The development of safety in these systems is significant and it has an impact on the new risks that are accepted in order to keep it at a stable and very low level.

However, low the remaining risk, it does not eliminate the possibility of an accident. A second paradox is that every accident that still occurs tends to be more severe due to the increased performance of the systems; it is sometimes incredibly expensive in terms of the compensation paid to victims, to the point where it results in a public crisis of confidence in many sectors.

The cost of compensation for a road accident involving a young person who was disabled for life in 2007 reached 7 million euro (source: MACSF [8]). It had been 10 to 20 times lower on average just 10 years previously, although road accidents were occurring twice as frequently.

The provision for a possible accident in the United States involving a Boeing 747 (source March, 2001) has risen from about $500 min 1995 to $750 min 2001, and is now flirting with $1,000 m. It is even higher for an Airbus A380. It should be noted that in air disasters, compensation has to be paid for two types of claims: injury to victims and economic damage. The attack on the World Trade Center (September 11th, 2001) was the perfect demonstration that the economic damage done by a single accident could be high enough (Wall Street was out of action for several weeks) to ruin insurers and reinsurers (several hundred billion US dollars); in the end the airline insurers did not pay; they negotiated (during the crisis) a very severe limitation on their threshold of compensation for collateral and economic damages. Putting it in clearer terms, governments have again—as in the past—become the insurers of their own major risks because the private market is no longer able to cover these risks.

These figures are almost insignificant in comparison with the 25,000 billion US dollars that the sub-prime crisis and the European debt crisis could cost the Western systems. Once again, governments have been forced to intervene on a massive scale to insure against risks that were supposed to be managed by a freely competitive market.

When the cycle ends, accidents become intolerable more because of their consequences than because of their frequency. Policies on regulation have to be changed as a result. This is not easy, however, since those policies have to shift:

- from a culture of accounting (frequency) for accidents and near-accidents (which are still dominant) which is easy to understand, put into practice and communicate to the general public, supported above all by the myth of zero accidents (accidents and incidents are counted and the reduction in numbers over time is highlighted);
- to a culture of justifying the limitations placed on investments in safety, an area with no upper limit (increasing investments that have no results either for the public or for the enterprises, in the absence of an accident, other than an increase in costs) and which is clearly less beneficial in terms of communication with the general public.

It can be imagined that the shift in policy that is required by the process of technical progress and growing intolerance of the remaining risks is difficult: on the one hand there are greater benefits and better objective safety, and on the other hand there are rare but serious accidents that result in an over-reaction from public opinion which has the potential to sweep away politicians and sometimes the industry itself with a complete lack of technical and economic rationality (industries flee to other countries, jobs are lost and confidence is lost in public figures who do not really have any control over the system).

Today this tipping point has not yet been reached in most sectors, and politicians in Western countries are continuing to develop regulations in reaction to events, on a case by case and sector by sector basis; ultimately this policy acts as a brake on technological escape which has to be applied from an increasing distance, and it therefore cannot claim to be able to offer a controlled exit from the societal trap in which Western countries now find themselves in terms of the management of their risk industry/practices.

To summarise, a societal transformation that is common to all European countries is leading to common problems in terms of risk management.

Alongside the Common Features, of Which There are Many in the Field of Safety, a Number of Macro-Scale Cultural and Strategic Differences Still Remain in Terms of Safety Interventions

All the observations made above converge on a common global model for the management of safety in large high-risk systems.

There are, however, a number of regional differences between micro-scale strategies in the area of safety, most of which are associated with cultural and legal models. One of the most obvious differences is between countries whose legal and administrative systems are inspired by Roman law (France, Italy, Greece and Spain) and countries with an Anglo-Saxon legal system (typically the United States and the United Kingdom). The differences are primarily expressed in terms of the scope and prescriptive force of actions in the area of safety.[5] One specific feature of Roman law, which was introduced by the emperor Justinian, is that it is based on written specifications. France, through its Napoleonic Code, is fully in line with this heritage of a written legal framework, as a country with a growing body of texts setting out the legal rules down to the last detail, including those governing aspects relating to culture,[6] as a "land of written laws", while systems in the Anglo-Saxon world create more room for laws based on precedent. Of course intermediate versions have gradually blurred the differences,[7]

The two examples below, drawn from the world of aviation, illustrate the tenor of these differences.

Roman versus Anglo-Saxon legal systems: different sensitivities. Example of the national regulations on medical examinations for airline pilots before European harmonisation
Until 2009, regulations existed at the national level.

[5] http://www.answers.com/topic/roman-law.

[6] Written reason of the Corpus Juris Civilis.

[7] For example, read: MacQueen [9].

In Sweden, these examinations are carried out by certified doctors who are registered in primary care as independent professionals. These doctors send their results to a national agency. The level of regulation is low but there is zero tolerance: there is a code of practice, continuous assessment of approved doctors with a reputation for being strict, and certificates are issued for a limited period (every accident involving a pilot results in a detailed examination of the aptitude files; any failure by a doctor to detect a direct or indirect cause of the accident may constitute grounds for loss or non-renewal of his certified status).

In France, the exclusive right to carry out professional medical examinations in the aviation industry has for a long time been reserved entirely for military examination centres. There was a tentative move in the late 1980s to include civil centres as well. Until recently, however, no registered, independent doctors in primary care were able to obtain certification. This centralised, public system has a more extensive and much stricter system of regulation (in theory) than the Swedish regulatory system (the consultations comprise a larger number of items and some aptitude standards are stricter); this strictness is offset by a pragmatic approach with multiple exceptions. There is virtually no continuous assessment of certified doctors. Certificates are issued for life, and the penalties in the case of errors are extremely marginal.

These differences were certainly reduced with the advent of the European regulations in 2000, but on closer inspection the cultural difference has not yet been eliminated.

Roman versus Anglo-Saxon legal systems: different sensitivities. Regulations for training crew members in the management of crew resources [crew resource management (CRM)]

The work that was done in the 1980s, mainly in the United States, showed the importance of training in what were called "non-technical" competencies and are often covered by the acronym CRM, for training of pilots working in air crews.

In 1993, the International Civil Aviation Organisation altered its own regulations and introduced a paragraph 9.3 into its Annex 6 (on aviation operations) which requires continuing training of crew members in non-technical skills.

From 1993 onwards, the French Direction Générale de l'Aviation Civile (DGAC) has had a document that requires airlines to develop CRM training courses (Decree of March 1993). This text is ambitious and highly prescriptive, with quite a short deadline for compliance. Administrative supervision is the responsibility of Regional Civil Aviation Directorates (directions régionales de l'Aviation civile—DRACs). Air France is currently the only airline with a compliant training tool (since 1992). Responsibility for training personnel in the regional Civil Aviation Directorates, guidance and certification of compliance is the responsibility of a small number of civil servants. A study carried out in 2006 [10] showed that this process of certifying compliance in French airlines is erratic, very heterogeneous from one company to another and that follow-up by regional administration agencies is very tolerant.

Conversely, since 1988 the FAA (US Federal Aviation Administration) has had an AC (advisory circular) recommending (not requiring) the practice of CRM by its airlines. The first airline in the world to have such a tool was United Airlines in 1982. In 1993, more than 70 % of American airlines had a CRM and there was a degree of national homogeneity. No formal regulations were put in place, however, until 2001.

Similarly, no formal regulatory requirement existed in the United Kingdom until 1999, which was the date of the changeover to the harmonised JAA European regulation. The research agency of the UK's CAA (UK Civil Aviation Authority) created financial incentives for spontaneous adoption of these techniques by companies after 1990, with assistance from and under supervision by the Royal Aeronautical Society. Most of the airlines complied before 2000, following a minimal but relatively homogeneous model.

These two cases, and other examples could easily be mentioned from both land and air transport (e.g. operator working times), road traffic regulations, the environment (chemicals, soil emissions), construction and certification of airframes, emphasise consistent cultural traits that create differences between ways of regulating safety in different regions of Europe.

In France, and more generally in the countries of Southern Europe (Italy, Greece, Spain and Portugal), the following are characteristic:

- regulations on safety issues tend to be viewed as an incentive for compliance by the industry or the public services being addressed. Rules are promulgated in a context where it is known that the industry—or public service—will not yet be able to comply with the regulation;
- this strategy leads to a large number of highly prescriptive regulations which are often issued at the highest levels of government. Conversely, few straight forward recommendations are made and these elicit little response from the industry, while little help or support for their implementation is provided by the State;
- compliance is understandably slow and many exceptions are made; these exceptions often become permanent ways of regulating the system;
- observatories and compliance tools are created, primarily in order to create a flow of information back to those issuing directives at the government level. They, on the other hand, do not feed back much information to production units lower down;

- the centralised nature of decision-making and monitoring tools reduces the effect of individual responsibility and instead promotes central responsibility. It is most commonly the State and the major organisations that pay out compensation in case of errors.
- The Anglo-Saxon bloc, typically the United States and the United Kingdom, tends to take the opposite approach:
- no rule unless the majority of the industry is capable of following it.
- on the other hand, a large number of recommendations are issued and these are closely followed up by the State and by associations. They meet with a real response from the industry (a recommendation issued by the FAA in the form of an AC [Advisory Circular] is practically equivalent to a rule in France in terms of the comparable effect on the industry);
- there are very few observatories and individual sanctions are more severe; the insurance rationale is different;
- Germans and Nordic countries occupy an intermediate position, with greater regionalism in terms of specific aspects of regulation, a system which is generally less demanding than the French system, but very low tolerance for departures from it.

What Lessons Can Be Drawn From This?

In every case, and contrary to the principles of common sense seen in the introduction, enhancing safety always has a paradoxical effect: it is a preoccupation of systems and societies that have already become safe; to some extent it is a preoccupation of the rich, once all their other essential needs have been met.

Worse still, the level of safety has the surprising property that it is never high enough, and it even leads to societal demands increasing in parallel with improvements in progress. It is a variable with no maximum, and all improvements lead to increases in the severity of judgements made by outsiders on the few breakdowns in safety that still occur. The more safety is required, the more it becomes an aspect which is difficult to manage and begins to seriously hamper the productivity of the system.

It is this “impossible safety”, together with a “brake on productivity” when people try to impose even greater safety, that creates the conditions for the emergence of a radical change that will allow a new life-cycle to begin based on a different innovation/industrial organisation.

In order to continue to exist and survive in an industrial paradigm which is nearing the end of its cycle, three types of action have to be taken which can be demonstrated to the general public:

- avoid going beyond what is demanded by society or moving too quickly towards ultra-safety (to preserve the remaining scope to make further progress);
- adjust the compromises between safety and performance and the necessary investments in accordance with the life cycle of the system. The appropriate safety tools and safety solutions change throughout the life cycle; this is one of

the most difficult aspects for the industry, since solutions that were still appropriate and effective recently may prove dangerous and inappropriate in the same system at a different phase in the industrial life cycle. In other words, the treatment for influenza in an older person with multiple comorbidities is not the same as the treatment for influenza in a young man of 20. The compromise in terms of the strength of treatment needs to be far more sophisticated in the case of the older person, since otherwise one may kill the person in the process of curing the influenza. The same idea of a compromise treatment is all the more important in the area of safety if one admits that the preoccupation with safety is a feature of ageing in the technological cycle;

- anticipating the crisis at the end of the cycle and investing in the next cycle in time to prevent one's own death at the end of the cycle. Once again, the need will not go away and the cycle will be followed by another cycle, but the industrial beneficiaries may change (and this may occur frequently). For example, the collapse of silver-based photography over a period of just a few months in the 2000s is an exaggerated example of the end of a cycle in which the industrial cards are suddenly reshuffled.

The following chapters will address the process of building the necessary safety compromises involving operators, teams (Chap. 2) and organisations and large systems (Chap. 3).

References

1. de Kersasdoué J (2007) Les prêcheurs de l'apocalypse, pour en finir avec les délires écologiques et sanitaires. Paris: Plon
2. Reason J (1990) Human error. Cambridge University Press, Cambridge
3. Hughes T (1983) Networks of power: electrification in Western countries. John Hopkins University Press, Baltimore
4. Amalberti R (2006) Optimum system safety and optimum system resilience: agonist or antagonist concepts? In: Hollnagel E, Woods D, Levison N. Resilience engineering: concepts and precepts, Aldershot, England: Ashgate: 238–256
5. Hale A, Swuste A (1998) Safety rules: procedural freedom or action constraint? Saf Sci 29:63–177

6 Rhode E (2007) Trop de loi tue la loi, la jungle législative, Le Monde, 24 janvier 2007:20

7. Amalberti R (2009) Quel futur et quelle stratégie de sécurité pour un système devenu ultrasûr ? Transfus Clin Biol 16:80–85
8. Robert S (2011) Macif: une indemnisation record pour un accident de la route, Published Dec 28, site MACIF, 2011, accessed on http://www.assurland.com/assureurs/actualite-macif/macif-uneindemnisation-record-pour-un-accident-de-la-route_18125.html
9. McQueen H (2000) Scots law and the road to the new ius commune. Electronic journal of comparative law, vol 4.4, December 2000, Accessed http://law.kub.nl/ejcl/44/art44-1.html
10. Deharvengt S (2006) Barriers to regulating resilience: example of pilots' crew resource management training. Accessed at http://www.ep.liu.se/ecp/023/002/ecp2307002.pdf

Chapter 2
Human Error at the Centre of the Debate on Safety

Abstract This chapter focuses on the discovery of the safety model that is used by individuals in order to carry out their work without incidents or accidents. It offers a micro-level perspective on safety. The material here is a summary of a book on the management of high-risk systems published in 1996, with additional insights from the latest work on common forms of bias in error analysis and the importance of the concepts of adequacy, compromise, trade-off and the central role of routines.

Human Errors, Major Steps Towards Building Knowledge

Human beings do not seek to work without making errors; they seek to achieve a satisfactory result while minimising negative costs (time wasted, incidents). A person's key objective is to make progress towards the goal while remaining cognitively in control of the situation. There are two aspects to this type of cognitive supervision: one monitors the progress made towards the goal, while checking the external results of what has been done, while the other is focused on keeping the cost of the cognitive performance of the work down to a reasonable level (fatigue, investment, sacrifice of other activities that could be done in parallel). In this context, the error flow is high (particularly routine errors) but (1) the error flow does not predict the risk of an accident and (2) the error flow must be seen in conjunction with another flow: the error detection and recovery flow, since the impairment of this flow is a better predictor of the risk of an accident.

It is widely known that 70 % of accidents have a human cause related to operator errors. Equally, if one adds to this the contribution of designers and managers to what are called technical errors (breakdowns) or organisational errors (management decisions, social climate), 100 % of accidents actually have direct or indirect causes associated with human factors.

Due to these figures, it is a natural priority to understand human errors in order to reduce them, while common sense suggests that reducing errors will necessarily lead to a reduction in accidents.

The reality is far from being as simple as that. This chapter sets out four observations: (1) errors occur even more frequently than people think, at a rate of several per hour, (2) but these are largely self-detected and recovered, so that the

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology, DOI: 10.1007/978-94-007-6549-8_2, © The Author(s) 2013

observed consequences are much lower than would be predicted from the error frequency, (3) they are inherent in cognitive function, particularly when it is routine, and they therefore cannot be eliminated except by eliminating human beings, (4), excessive—and erroneous—simplification of the link between errors and safety has not really resolved the questions of safety.

Systems that are designed on the basis of contradictions and built on weak scientific foundations do not allow operators to engage with them effectively. They result in a vicious circle, simply shifting errors elsewhere and making them more difficult to control and manage. Pursuing this rationale of course leads to the introduction of more computerisation in order to (finally) obtain true reliability. This involves pitting human reliability against technical reliability, which results in utter failure to achieve synergies or summative effects between the two. The results are inevitably worse than expected.

A more appropriate approach would be to analyse this link between the error and the accident, to pass through the mirror, see through the operator's eyes and understand that the management of individual risks is based on extremely sophisticated knowledge of compromises and overall control of the situation. The error itself never causes the risk of an accident; it is losing control, losing awareness of the compromises between acceptable risks and losing the ability to manage the situation that can very quickly lead to an accident.

That is why this chapter on errors and the management of individual risks incorporates quite a solid theoretical framework. Here, even more than in subsequent chapters, there are some false "good ideas" that need to be corrected.

It was only recently, in the 1970s, that the study of human error became an object of separate scientific study for psychologists. Prior to this, with the exception of the Gestaltists in the first half of the 20th century, errors were seen as just one of the many performance scores in the experimental approach to physical or psychological phenomena.

The Initial Contribution from Gestalt Theory was that Failure Makes it Possible to Achieve Understanding

The first significant work on errors (more specifically on failure) was done before the war (during the period from 1910 to 1940) and is classified under Gestalt theory or the theory of forms. This theory is considered to be the foundation of modern cognitive psychology.

The Gestaltists (Koffka, Köhler and Wertheimer) were primarily interested in the organisation of the visual environment that requires our brains to make what are sometimes incorrect perceptual interpretations of complex scenes (ambiguous shapes).

Everyone has come across these complex shapes that give rise to illusions of interpretation.

One example of an illusion of interpretation described by the Gestaltists. This variant of the Müeller-Lyer Illusion uses two arrows. When asked to compare the

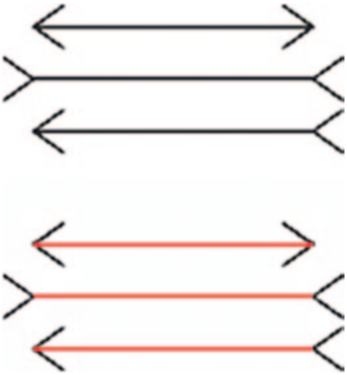


Fig. 2.1 The observer states that the line with the points facing towards the middle is longer

length of the lines (excluding the arrowheads), which are actually equal, the observer states that the line with the points facing towards the middle is longer (Fig. 2.1).

This approach to perception very quickly leads to the realisation that seeing is not a simple objective process (not everyone sees the same thing in the same situation). It is more an active process of construction by our own cognitive apparatus (guided by our knowledge and expectations), filtering and correcting the properties of the environment to read in it what one is looking for. This active process of construction, which has emerged from the work done on perception, was quickly extended to theories in the social domain in the 1930s [1], and then to the understanding of complex situations and decision-making in the 1940s [2].

It is the reorganisation of premises (the initial conditions for reasoning and initial perception of the situation) in order to be able to modify one's cognitive field (think of another possible solution) that leads to a reconsideration of the observable facts (seeing new things that have not been seen until that point) and finally allows the solution to stand out (insight).

When the subject makes his decision while adhering to his initial impression, he generally reproduces a known solution. Duncker shows that faster, more elegant solutions often exist which are not even conceived by the subject as long as his routine solution works. It is only when he faces failure that the operator reviews his hypotheses, reconsiders the facts that are available in the situation and produces (rather than simply reproducing) a solution.

For the Gestaltists, failure and reaching an impasse are an important or even indispensable condition for unlocking understanding and producing new ideas. This positive view of failure was to feed a large proportion of the modern literature on errors.

The First Works on Error: The Essential Role of the Control of Cognitive Activity

The second starting-point in the study of error emerged more recently and in a completely different way, since it is an extension of the debate on theories of attention and routines.

The first models of attention solely emphasised limitations. The limited channel model of Broadbent [3] interpreted attention as a filter prioritising the information available in the outside world and allowing it to penetrate cognition via a pipeline or “single channel”, following an order of priority. Miller [4] confirmed this constraint by showing that short-term memory was limited in duration and capacity (7 elements ± 1).

These approaches were quickly criticized for their failure to match the reality, since data quickly accumulated to show that the predicted limits were easy for any operator to exceed. Shiffrin and Schneider [5] therefore put forward a model, which has become famous and which identified two separate levels working in parallel, with interactive loops between them:

- a conscious, controlled level, requiring attention-based control. This level is limited in volume and duration (a driver who needs to devote his attention to finding a way out of a complex junction will stop talking for a few moments and miss what is being said on the radio, because he does not have enough resources to do two things at once);
- an automatic, routine level, not under attention-based control and with virtually unlimited parallel capacity (although the driver has stopped the conversation when entering the junction, he continues to be able to drive the car, using routines for changing speed and braking which are not really conscious, and he remains capable of managing a large number of low-level activities in parallel, such as operating the volume control, using the indicators etc.).

These ideas would then be put into operation in terms of a workload, using the metaphor of cognition with a reservoir of available resources [6]. Processes requiring attention consume resources, while routine processes do not draw on this reservoir. Experts have a better ability than beginners to use their routines and manage this reservoir, which gives them a greater ability to manage situations with a high workload.

During the course of this work, Donald Norman was the first author to use these ideas and deduct a theory from them on errors in routines by pointing out the paradox that this mostly affects experts.

The first model that he suggested [7] comprises two dimensions:

- a horizontal dimension containing a series of threads that function independently; each thread works according to well-known, routine procedures (cognitive psychology refers to these procedures as schemas or scripts);
- a vertical dimension that interacts with the horizontal structure to guide and regulate it.

The horizontal level makes it possible to carry out routine activities without control, as long as the action is progressing normally towards its goal. Attention and motivation take the form of vertical variables which modulate the activation of these threads (schemas) whenever obstacles or moments of saturation are encountered or when choices have to be made between current goals and routines.

Norman [8] then deduces multiple methods by which management of these routine schemas breaks down. These errors are referred to as *slips*. He identifies:

- slips resulting from incorrect activation of schemas: this may be an involuntary activation ("go too close to a well-established habit and it will capture your behaviour". For example, if you have to make an unusual detour in order to pick someone up, at the first junction which you take every day, you forget your rendezvous and find yourself back in front of your house [routine capture]). The schema can also lose its relevance: it may continue to function even after the person has forgotten why they started the activity;
- slips resulting from incorrect initiation of schemas. The schema is chosen and activated correctly, but at the wrong time, or it is mixed up with another schema and the result is incorrect: a secretary is typing a letter while thinking about her appointment at 12.30 and writes "the meeting will be held at 12.30" on the letter rather than writing the correct words: "the meeting will be held at 14.00". This may also involve a change in the sequence involved in executing a macro-routine which ultimately results in part of the work that needs to be done being skipped or forgotten: a person waters all the plants in the lounge every morning after getting up, but on a certain day there are friends sleeping in the lounge and it is not possible to go in. The task of watering the plants is put off until later, and the person ends up forgetting it.

The Contribution Made by Rasmussen: The SRK Model

In 1983,[1] Rasmussen [9] introduced the celebrated SRK model, which identifies three modes of cognitive functioning and three types of errors. He identifies:

- a level based on knowledge (*knowledge-based behaviour*): mobilising everything a person knows to understand the situation and take action in it, following a rational process, typical of processes learned at school;
- a level based on rules (*rule-based behaviour*): professional rules (if, then) are mobilised, making it possible to achieve greater pragmatism and more effective action than in the previous mode. For example, consider a simple cooking rule: "only put the pasta into the water once it is boiling"; there is no need to relearn,

[1] Jens Rasmussen is a visionary engineer, self-taught in human factors, who is capable of reading and bridging different streams of theory that are mutually unaware of each other. He refocused his career on technical and human reliability after the nuclear accident which occurred at Three Mile Island in the United States in March 1979. He was to become one of the pioneers of modern approaches to safety in complex systems and went on to have a profound influence on a whole generation of researchers who studied directly under him, such as James Reason, Erik Hollnagel, Dave Woods, and… the author of this book.

before putting the pasta into the water, why it is necessary to wait for it to boil, why the water is boiling, why water evaporates when it is boiling and why its boiling temperature changes with altitude…. Knowledge of the rule makes it possible to act effectively without asking “why”;

- the level based on routines (*skill-based behaviour*). The action becomes completely automatic in response to the stimulus: I see my house and I begin to take out my keys without even being aware of it.

This distinction has been part of the context of human reliability right from the beginning.

All learning starts from a way of working that is based on knowledge and ends with a way of working that is based on habits and routines. What characterises an expert is the increased availability of these routines, allowing him to work more quickly and cope with a larger workload.

Routines are thus markers of expertise above all and form the habitual basis on which a professional works. Only when routines fail does it become increasingly necessary to revert to other methods, which is both costly and hazardous in terms of the cognitive load. If the routine no longer works and progress towards the goal is blocked, the operator will shift to a rule-based mode and if he does not find a rule to rescue him, he will switch to a mode that is based on all his knowledge.

For example, you leave home to meet someone and you think you know the way. You are driving in a routine way, listening to the radio and thinking about your meeting. However, you turn right too early…. You have to shift out of routine driving mode (stop listening to the radio and concentrate on finding the way) and you will very probably try to mobilise a rule in your memory that might help you; for example:

“if I have turned off too soon, I should turn round and go back, unless there is a traffic jam in the opposite direction “or” if I have turned off too soon, I must be travelling in parallel and if I carry on I only need to turn left and then right again and I will be back on the right road”. So you try out one of these rules. If the situation does not improve, you admit that you are “lost” and no doubt switch to functioning in a much more analytical way, based on all your general knowledge: getting out a road map, looking for a street plan or asking for help. Each of these steps creates opportunities for different errors (routine errors, rule errors or knowledge errors).

To summarise, inter- and intra-operator performance variability (on repetitive tasks) is largely related to variations in this level of control over cognitive activity. Under normal circumstances, skilled operators make maximum use of the level based on habits (routines), and the cost of this approach is that they make a large number of routine errors. When the situation becomes less familiar, the subjects switch to a more attentive form of control and follow the rules more formally, or in the worst situations, they create new procedures from scratch; at that point their errors will most often be rule errors and knowledge errors.

The Summary by James Reason

Reason [10][2], inspired by Rasmussen's SRK model, again addressed this classification of errors into three categories, which is still the most authoritative classification:

- routine errors corresponding to functioning based on Rasmussen's routines (*skill-based behaviour*). These are errors in monitoring the work as it is done. The action is carried out without conscious control, in the context of a familiar type of work. The subject has not become aware that he has encountered a problem. These errors are characteristic of work done by highly trained experts. They are numerous (80 % of the total errors made) but a very large proportion are recovered (90 %) and, contrary to what is often said, they rarely give rise to serious accidents (but they are often responsible for incidents and oversights);
- errors of rule activation. The subject encounters a difficulty which he cannot resolve in a routine way (he is aware that he has a problem). The error will result from choosing the wrong solution by activating the wrong rule. This type of error does not mean that the subject does not have knowledge of the correct solution; he has not, however, been able to activate it, recover it from his memory, or (due to lack of time) he has not been able to use it in his situation. Another solution, which is less valid but is immediately available, has prevailed in his chosen approach. These errors are less frequent (15 % of all errors) but they are feared more than routine errors because of their consequences in terms of safety. They are often called "errors of representation" because the operator "applies his procedure correctly, but in the wrong context, where the procedure is not relevant". The problem of "fixation errors" (not changing one's view, becoming fixed in an incorrect perspective) is a specific subset of these errors. This group of errors are frequently addressed in the literature because they are difficult to resolve. The safe solution appears to be based more on working well as a team and the ability to adopt different view of the problem in real time [11, 12];
- errors due to lack of knowledge.[3] The subject does not know the solution to the problem that he has to solve. He mobilises all his cognitive ability, slowly and step by step, to come up with a new solution. The error may then take different forms: the right solution but too late, the wrong solution etc. This type of error is (fortunately) rare among professionals (less than 5 % of total errors) but it is clearly always more severe in terms of its safety consequences.

[2] James Reason was Professor at the University of Manchester for many years, and is now retired; he is no doubt the best-known theoretical author on human error. He has published a number of works, one of which is the reference work on this subject; he was strongly influenced by Jens Rasmussen, with whom he worked closely in the mid-1980s.

[3] When translating, one must be aware of false friends: these may be referred to as FAULTS in English, but the translation into French should not be FAUTE (which has too many connotations) but ERREUR DE CONNAISSANCE.

Differents types of errors and their characteristics (inspired by Reason [10])

Dimension	Errors based on automatic behaviour	Rule-based errors	Knowledge-based errors
Type of activity	Routine actions	Problem-solving activities	
Focus of attention	On something other than the task in hand	On considerations associated with the problem	
Control mode	Schemas	Stored rules	Limited conscious processes
Predictable nature of the error	Largely predictable	Variable	
Frequency	High in absolute terms, but paradoxically low compared with the large number of routines	Low in absolute terms, but high compared with the very small number of situations involving virtually total lack of knowledge	
Capacity for detection	High	Very low without outside intervention	
Risks to safety	Moderate	High to very high	

Work on Detection and Recovery

Discussing a mechanism by which an error is produced does not constitute an analysis of the error, far from it. The error only becomes a problem as a result of its consequences. Early error detection and recovery from errors before they have consequences form the very heart of risk management.

These types of error detection and error recovery are particularly effective in humans.

Hayes and Flower [13] were the first to take an interest in the ability of editors to detect spelling and syntax errors. They identified two separate mechanisms: (1) intentional detection when re-reading (editing) and (2) iterative detection during writing (reviewing), which is far more effective.

Allwood and Montgomery [14] added to these early works and supplied a theoretical context, based on work done on errors made by students in physics and mathematics exercises. They identified three separate phases in the correction process: detection, problem diagnosis and recovery. Detection simply meant perceiving that there was a problem during the course of the action (without identifying it). Diagnosis meant identifying the error. Recovery meant eliminating the error itself or its consequences.

These early studies concluded that there are four families of error detection strategies:

- strategy 1: types of evaluation based on knowledge about the result (*affirmative evaluation*). The subject checks his result on the basis of realistic ranges that he knows should encompass the expected result;
- strategy 2: routine checks (*standard check*). The subject carries out a check independently of any specific suspicion, and discovers his error;

- strategy 3: focused checks (*direct-error-hypothesis formation*). The subject responds to a bizarre result and immediately forms a hypothesis on the type of error that he may have committed;
- strategy 4: simple suspicion (*error-suspicion*). Part of the result is considered to be bizarre but it is not possible to formulate an explanatory hypothesis.

The strategy that detects the largest number of errors by volume is "direct-error-hypothesis formation", followed in order of effectiveness by "error-suspicion", "affirmative evaluation" and a long way behind: "standard check", which corresponds to the type of checking methods learned at school.

In sum, these strategies are impressively effective.

A total of 70–80 % of the errors that are made are detected by the person who committed them within a very short time: 90 % of routine errors and, not surprisingly, only 20 % of knowledge errors [15–17].

These works also teach us that the best subjects carry out more standard checks, although as we have just seen this strategy is apparently not very efficient in terms of detecting errors. No doubt the errors that it does detect cannot be identified by the other strategies and it is this fact that makes the difference between subjects who fail to carry out these systematic checks and experts.

More importantly still, Allwood (op cit.) shows:

- that efficiency in solving a problem is significantly correlated with the proportion of errors detected when solving it;
- that there is no correlation between the number of errors made and the subject's ultimate effectiveness.

Combating good ideas that are nevertheless incorrect: **the volume of errors does not predict performance;** it is error recovery that is the best predictor of the subject's performance.

Errors that are made seem to help the subject to be aware of his activity and control the process of making cognitive compromises in order to converge on a solution.

The subject uses the errors that he makes to engage in continuous self-evaluation of his cognitive function and control his risk-taking. Reflective activities (watching oneself work) are clearly central to this control process.

These Results have been Validated in Industrial Situations

One of the first industrial applications [18] involved a situation at a printing press where a database management system contained a large number of tasks that had to be managed in parallel and tasks whose complexity varied from one workplace to another. The study showed that **the number of slips increases as the task becomes more complex, but the number of slips detected also increases as the subjects become more experienced**. The study confirmed that it was routine checking processes that contributed to this significant improvement in performance among expert subjects.

The same study showed that rule-based errors were no more frequent when the task was more complex and that these are not detected significantly more effectively by those with more experience. Rare errors of knowledge, however, are detected much better by expert subjects. This work provides spectacular confirmation of the complexity of managing cognitive compromises.

Combating good ideas that are nevertheless incorrect: **more routines are used when the task becomes more complex.** When the task becomes more complex there are more areas that are not understood. The subject is worried about committing errors of understanding and makes it a priority to invest his resources in activities related to understanding, to the detriment of activities that he believes he has mastered, which he then completes routinely and without checking. The paradox, of course, is that he makes more and more routine-based errors.

The expert is concerned about these routine-based errors and protects himself against them by using serial checks. Ultimately it is these routine-based errors that he makes most frequently, simply because the limitation of resources forces him to make the greatest possible use of automated behaviours. This clearly reveals the deep defences of the cognitive system, which has no choice at the outset other than to take risks (by automating behaviours) in order to cope with the temporary shortage of available resources, but then protects itself from the risks that have been taken by using a series of checks.

Combating good ideas that are nevertheless incorrect: the spontaneous error rate is high among humans, but does not predict accidents. This rate may reach 10 errors per hour under inattentive, relaxed conditions.

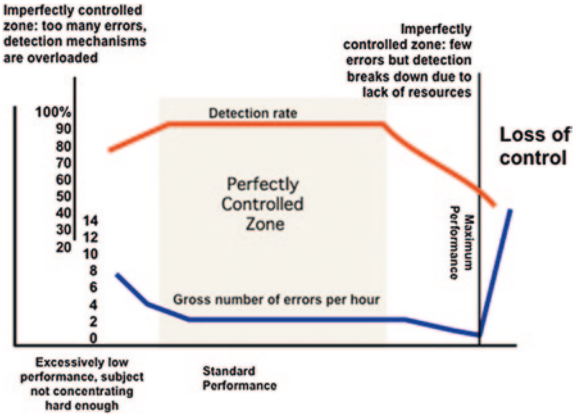
Under more attentive conditions, the average error rate is closer to two errors per hour (the rate observed in civil aviation over a series of more than 3,000 flights; these results were obtained from the Line Oriented Safety Audit (LOSA) type of large-scale online audit techniques [19, 20]).

These error flows do not predict many accidents, since the vast majority of such errors, if not all of them, are detected and recovered by the operator himself.

When the situation requires greater attention with greater challenges in terms of performance, the operator can reduce his error rate still further to about 0.5 errors per hour. Paradoxically, however, an operator will lose control in these extreme situations not because he is making more errors (he makes fewer), but because his system of control gets out of balance and he no longer has sufficient resources to recover from the few errors that are still made.

The accumulation of results on how errors are made and detected [21] leads us to consider these two phenomena as linked within a single cognitive approach. Human reliability is based on a system in dynamic equilibrium, in which an error generation rate is linked to a detection and recovery rate.

The system breaks down at both extremes of performance, either because the subject is not concentrating sufficiently hard and the error rate ends up exceeding the detection rate, or because the subject is concentrating too hard and commits few errors but in the process consumes all the resources needed for the automatic cognitive detection feedback loops. In the latter case, loss of control paradoxically occurs at a time when the subject is committing almost no errors at all (De Keyser [22]; Wioland, Amalberti, op. cit.).



In reality it must also be noted that a significant proportion of errors that are detected do not have to be recovered by returning to the error and immediately correcting it (UNDO), simply because many of these errors (1) have no immediate consequences (leaving the light on in the office), or (2) create new options which are just as acceptable as those imagined before the error was made (you intended to go along a certain street but missed the turning, reorganised your plan, did not undo the initial error and adopted a new itinerary which is still compatible with the intended destination).

All the production, detection and recovery mechanisms are covered by the term ***error management***.

Also in medicine…. The safest hospitals are not those where no more errors are made, but those that detect and recover from the errors that they have made most effectively [23].

The authors studied the adjusted hospital mortality for a cohort of 84,730 patients who had undergone vascular surgery procedures throughout the United States. The mortality rate varied considerably from one centre to another (3.9–6.9 %) and the variable that best accounted for the mortality rate was not the rate of benign or severe complications that occurred in these hospitals (which was virtually constant for all institutions) but the incorrect management of those complications. Patients in the hospitals with high mortality rates were twice as likely to die from their major complication as those in the safest hospitals. This significant result supports the increasingly widespread idea that the traditional approach to patient safety fails to address a number of vital aspects of risk control because it focuses too much on preventing and avoiding problems and not enough on recovering from problems that have already occurred. There is a need to carry out a true reappraisal and repositioning of the approach to safety.

Three Recurrent Biases in Relation to Human Error

The study of human error is riddled with bias. Three forms of bias are particularly significant: reconstruction after a setback, excessive attribution of error causality to front-line operators, and inaccurate links between errors and accidents.

Industrial disasters have played a major part in generating the fascination for studying human faults and human errors. Without the nuclear industry and its disasters at Three Mile Island and Chernobyl, and more recently at Fukushima, without the aviation industry and the accident in Tenerife, and without the Bhopal disaster in the chemical industry, very little progress would have been made towards theories on error, safety and human reliability.

On the other hand, the knowledge of errors that is available has been used particularly intensively in work on safety in complex systems, but this has not always been successful due to the many contradictions or imperfections that arise when making the transition from theory to practice.

Despite, or because of, this profusion of fashionable literature in which any assertion can be supported or contradicted, there are three recurring types of bias that have become established in the use of accident analysis in industry.

Hindsight

The first bias is that of post hoc reconstruction or "**hindsight**" in relation to the history of the accident. There is a temptation to assume that the operator behaves rationally and pays attention to everything, and to judge him on the basis of what has been discovered during the investigation, particularly previous tell-tale incidents that should have alerted him. In most cases, however, the operator was working in a routine way, was not aware of the previous tell-tale incidents and did not imagine that he was exposing himself to disastrous conditions through his decisions. All deviations from an idealised form of adherence to procedure are seen in hindsight as errors or violations, while in reality such deviations are justified by the reality of the context at the time (management of the moment-by-moment workload, anticipation, external disturbances etc.[4]).

Attributing all the Blame to the Last Person Who Carried Out the Action Causing the Disaster

The second form of bias is **excessive attribution of the causation of accidents to front-line operators** (the people who are involved in the action). Factors associated with the overall complexity of the system are lost in most of our rational analyses which are intended to break down the work into distinct components [24]. NB: this does not mean simply switching the analysis by considering that causes that are excessively operator-centred should now be analysed in a "deep" way, placing the blame on latent errors of design and organisation. Reason's Swiss Cheese model has unfortunately given rise to this bias in industry which is just as serious as the earlier types… because it only shifts the "token" cause to one or other of these parties and always ends up committing the same errors in terms of attributing blame and fails to address "the whole" (on this subject read the cited work by Dekker op.ci., or Johnson and Holloway [25]). On the contrary, the challenge is to consider the model of dynamic linkages between all the parts in the system.

A third bias tends in the same direction: all too often the analysis is limited to considering the usual range of causes that are already known and catalogued (operators, organisation, management, design). Chris Johnson speaks of a "lack of imagination" [26] among analysts who are "incapable of seeing the non-standard" if it is possible to blame one of the usual causes. The result is disastrous both in terms of the understanding of accidents and in terms of the action that is taken following accidents: the decisions that are made to take action in relation to each sub-sector that is judged to be at fault lead to growing complexity in local protection systems. These are often mutually contradictory because they are designed in isolation from each other and the results are at best ineffective and at worst more dangerous in terms of the overall situation. Fortunately, or perhaps unfortunately, this safety handicap associated with the absence of an overall vision is only a characteristic of the safest industrial systems; simpler systems have long benefited from local action. A safety model which is managed piecemeal and at an excessively local level only really becomes a problem once this system is made safe, but it is also more difficult to abandon because it is supported by the memory of all the past successes as the system has progressed.

[4] Read the very good commentary by Dave Woods on hindsight bias in the enquiry on the Columbia shuttle accident http://researchnews.osu.edu/archive/hindbias.htm.

Confounding the Error and the Accident

As a result of focusing on errors, permanent ambiguities have become established in relation to the link between errors and accidents. The two terms are often confounded, and all errors are demonised in the quest for an optimised cognitive system that works in a similar way to a machine.

The structural role of errors in solving problems has been minimised for twenty years. The accumulating evidence that operators make a large number of errors but recover the majority of them, has also been neglected.

It is also too easy to forget that making errors (particularly routine errors) is the price that is paid for working quickly, and consequently the price of a degree of social and economic efficiency. The price to pay for seeking to control everything and avoid all errors is usually such slowness in implementation that the most worrisome risk becomes that of "not doing the work at all".

A nurse could thus probably reduce the number of routine errors that she makes by concentrating on every individual job that she carries out, like a factory worker, but in that case she would probably treat five times fewer patients in a morning (certainly if she were a newly qualified nurse). If the criterion of systemic analysis is used, one can imagine that more patients might be put at risk by making no routine-based errors at all in the management of a patient transferred to the ward (routine-based errors which are, as we have already seen, 90 % recovered with no real consequences for the patient), while accepting the secondary risk of not treating patients who are transferred to the same ward since there is not enough time available.

It has therefore been necessary to await more favourable circumstances before starting the process of changing the dominant way of thinking about error, at least in the research domain. The industry has become aware of two recurrent problems in traditional approaches to safety: (a) the accident rate reached a plateau despite optimising solutions to block errors [16, op. cit.] and (b) the use of increasing numbers of procedures to reduce the number of incidents and accidents has sown the seeds of reduced adaptability on the part of operators, so that they have lost part of their ability to manage risks.

All the conditions were in place for a theoretical and practical shift in ideas about human reliability. In just a few years, the research landscape has changed and there has been in-depth revision of what is understood as "good" cognitive functioning.

"Good" cognitive functioning by operators, which is what enterprises are looking for when they seek to become "safer", should no longer be expressed in terms of seeking to work entirely without errors, and particularly not working with zero instantaneous waste (avoidance of all errors and faults or absolutely immediate recovery, with minimal response times and maximal understanding).

Instead, it takes the form of compromises that make it possible to achieve the goal (or one should really say "goals") in a dynamic way and at an adequate level of performance.

There are three key ideas in this theoretical revision:

- the idea of adequate performance, which is often wrongly understood as reflecting a certain type of laziness or laxness. Instead it should be understood as an appropriate response to the environment, which offers social satisfaction to the person doing the work, taking into account his goals, the context, the views of others, the expectations of society and what he is capable of doing. The concept of "adequacy" is considered separately for each type of work and is not in conflict with very high performance and high cognitive cost (for example on the day of an examination);
- the concept of dynamic adaptation, with significant fluctuations in performance over time but ultimately an acceptable response overall and within the intended time. The time available and the intended final outcomes are the units against which cognitive performance should be judged, not the results seen at every moment during periods before the deadlines are reached. In the end it turns out that errors are simply the price that has to be paid for a well-controlled compromise and that they are often only secondary variables in terms of keeping the situation under control;
- finally, the concept of metacognition or reflectiveness (looking at oneself) which makes it possible to control risk management in a way that is acceptable and accepted, and in particular the initial performance contract.

The Concept of "Adequacy" as a Cognitive Tool for Management of Contradictory Risks

> Maximum performance is almost never required of operators (and that performance would also differ considerably from one operator to another). What is required, however, is "adequate performance" in order to achieve the social objective of the production system (the performance that all operators can achieve and which is therefore more predictable). This concept of "adequacy" (sufficient action) forms a practical reflection of the intention at the time when the work is done and of our understanding of social expectations in general.
>
> It is applicable in the area of safety, as in other areas. Every operator constantly integrates and adjusts his representation of what is "adequate" in the context in which he finds himself.

The assessment of "adequacy" is based on very varied and highly sophisticated cognitive mechanisms, some of which are automated when managed expertly, and these can also explain deviations in terms of risk-taking in the absence of certain precautions in terms of how the workplace is organised.

Adequacy in Mental Representation and Planning

It is completely useless for a human being seeking to decide on a course of action to look for a perfect match between the real world and his own representation of it; indeed doing this would represents a disability.

A number of schools which are geographically remote from each other (Norman in the United States, [27]; Ochanine in Russia, [28]; Piaget in France, [29]) had all emphasised very early on, in different words, the distortion inherent in mental models as compared with the real world, their simplicity and emphasis on purpose and all *ultimately* stressed the usefulness of those distortions in terms of success in action (and communication).

These simplifications and distortions of the real world are responses to the psychological impossibility, in terms of both intensity and quality, of perceiving, knowing, understanding and doing everything (this idea is also at the centre of the work of Herbert Simon, who won the Nobel Prize in 1978 for his work on bounded rationality [30]).

In addition, the mental model[5] is not primarily intended to reflect something real, but its purpose is to predict what one is going to do and what will happen, and that function is essential.

The representation of the world allows the operator to mentally transform the world, to be concerned about events and anticipate corrections (Piaget's concept of the pre-correction function, op cit.). This is shown by the fact that doctors, who are experts in standardisation, and expert pilots, spend more time avoiding problems in advance than managing real problems [31]. Those working in the areas of planning and problem-solving have regularly identified these properties of adaptation and correction by anticipation [32].

Conversely, the benefit of planning and anticipating everything reaches its limits in the way work is done in practice by the operator. He actively plans alternatives as long as there is doubt concerning the credibility of his chosen solution, or if he considers that the cost of the resources he is using is too great, and particularly if it is not easy enough for him. Planning, however, often stops short of the maximum capacity for refinement of which it would be capable. This is referred to [33, 34] as the "useful cognitive cost": what would be the benefit to the operator of developing a more sophisticated plan, if it involves adding elements that will become obsolete (because the action is not carried out immediately and the context is going to change) or if what has already been achieved is robust enough, taking into account the knowledge available to him and the challenges that exist? In fact the most important plan to make before taking action should be above all to define the intended result (the performance contract), identify the probable points of difficulty in implementation and protect oneself against these or avoid them through prior reflection; the remainder of the implementation process can easily be done using in-line adaptation, as all the studies on combat pilot training and preparation for high-risk situations have shown [35, 36].

[5] The terms "mental model" and "mental representation" are synonymous and may be interchanged.

Adequacy in Decision-Making

The renewal of theories on natural decision-making [37, 38] has provided a number of arguments for the concept of adequacy. Those working in this area have gathered observational data from almost all high-risk industries (civil aviation, public transport, the nuclear industry, the chemical industry, firefighting, armed forces, emergency medical care etc.).

This provides evidence that the types of bias put forward by traditional theories on human decision-making [39] do not, in fact, have any real importance or relevance in natural, complex, dynamic situations.

Observations on the ground actually show that decision-making is a continuous process which is linked to the environment. This process takes the form of partial decisions, which are more or less relevant but do generally lead to acceptable results, taking account of the margins that exist in real situations. In many cases, decision-making processes in context are guided to some extent by the features of the situation (affordance[6]). Finally, operators often have a good knowledge of the “worlds” to which these decisions apply, so much so that decisions which are theoretically not very valid are ultimately not very dangerous, mainly thanks to appropriate responses by other cognitive agents nearby; for better or worse, operators have extensive expertise in relation to what they are able to control in terms of deviations and they will therefore tolerate a situation where their own decision is not very valid, as long as they think this will not lead them into a situation from which their expertise cannot extricate them.

Adequacy in the Areas of Control and Implementation

Taking into account the “adequacy” of its representation, the mental model certainly does not specify the whole procedure that must be put in place in order to do the work; it only includes essential guiding aspects and relies almost uniquely on routine interaction with markers that are read from the environment in order to make progress towards the goal.

Gibson and Crooks [40], in a historical article (1938) on driving automobiles, spoke about spontaneous attraction (by the affordance space). Affordance spaces are desirable areas that draw action towards them (safe field of travel: regions

[6] The term “affordance” is a neologism, originally coined in English, which means the idea of inciting or inviting action. It relates to a physical structure in the environment that spontaneously favours a specific action on that physical structure (for example pushing or pulling a door, depending on the shape of the handle (see Norman [41] for a development of the concept inspired by Gibson).

without obstacles, clearly lit, where the possibilities for interaction can be perceived) which emerge from the environment and are overlaid onto an automated perceptual guidance system, while dangerous areas are naturally avoided (areas in shadow or where the ground is poorly visible). This desirable space integrates the results of all perceived or imagined restrictions, including restrictions on one's own ability to act (this is certainly a forerunner of the concept of problem space put forward by Newell and Simon [42], except that here the space is mostly guided by the environment and its external, physical representations).

The model predicts that, as the operator reads the environment, he will seek out the most attractive path, which makes the most sense in order to progress in a routine way towards the destination. Maturana and Varela [43] go even further, considering that "perceived reality" is nothing but a consensual construction, born of this dynamic coupling with action to guide the actions that lie within a person's capability. These authors use the terms autopoiesis and enaction to indicate that the "perceived reality" literally emerges from the contact between motivations and local circumstances and that it is constantly evolving according to these dynamic links.

Two Levels of Supervision

All this work on adequacy converges towards a cognitive model of the individual, who is said to manage two types of supervision in parallel to ensure that he remains in control of the situation: management of the physical process and the situation (referred to as external control or supervision) and management of himself as a cognitive actor in the process (referred to as internal control or supervision). These two forms of supervision, whose interests often conflict, explain the fundamental need for compromise mechanisms and adequacy.

Supervision of the external process [44] permits intensive use of routines while relying on planning and guidance obtained from the affordance within the environment. It is only in situations where problems arise and routine processes are blocked that cognitive processes are invoked; in that case cognition has to be used more intensively and that intensity must also be controlled so that it produces results that are useful before the deadline for the process, which continues to evolve [45].

Orientation of roundsmen in the corridors of nuclear power stations.

A study using a realistic simulation [45] showed that roundsmen in a nuclear power station only check their own progress as they move around against a small number of key points and that they are not really aware of these checks. Sixty percent of those key points correspond to places where confusion between corridors is at a maximum (same colours, various similarities), creating an obvious source of errors. These checks are relatively common to all the roundsmen, as if they were triggered simply by contact with specific features in the environment. On the other hand, the remaining 40 % of checks are not linked to the environment, and are variable above all depending on the pressure of work for each roundsman. All this suggests that these controls are oriented mainly towards supervision of their own behaviour (I have allowed myself to become distracted, or I am thinking too much about this and not enough about my immediate work).

When these roundsmen are exposed to a change in time pressure (having to work more quickly), the external checks remain identical but the personal checks are organised differently, are closer together in time, probably in order to offer better monitoring of the automatically increased risk of errors and overload.

Internal or cognitive supervision (of the mental process) responds to objectives that complement the external supervision of the physical process [46]:

- on the one hand this means deciding when is the right time to initiate cognitive operations that require attention (and are therefore limited in number), focused on (re)planning, understanding and building new solutions, when routines are blocked or the self-evaluation of performance is negative;
- on the other hand, since processes requiring attention are slow and sequential, it is necessary to choose priorities continuously and frequently also to decide what interruptions (of the thought processes currently taking place) should be made in order to free up resources (reduce the mental burden) and to be able to think about priority items. Metacognition (looking at oneself) is deployed extensively in all these trade-offs. We are beginning to understand the nature of these trade-offs. Some solutions are local: the operator usually gives priority to completing current tasks before opening up other avenues for investment. The operator is also able to carry out very sophisticated checks to manage the sharing out of tasks in real time, as has been shown by the results of tests on airline pilots [47]. Pilots have the know-how to be able to shift from one task to another while minimising risks: estimating the time deadline, the time remaining before the previous task is completed, estimating stability and predicting tasks in the immediate future, using informal redundancy networks and symbolic referencing of warning signals to return to a task that has been left pending. Other solutions rely on opening parallel cognitive loops, which will work on the process at different levels of temporal depth. This parallelism, which necessarily results in sub-optimal performance for each specific activity (because it is necessary to divide one's attention) rarely proves disastrous because real situations are much more tolerant than laboratory situations. The low level of demand from the outside world generates scope to make actions effective and automatically reduces the effects of errors. This low level of demand is of course no accident: it results largely from the organisation of the world and of professions, in which people, shaping their environment, generate their own margins for action and their own "affordance".

As we have just seen, adequate action involves the use of time and levels of understanding.

Using Time to Control Adequacy

After adequacy, the time available is the second important variable which is poorly understood in the literature on reliability. Experimental approaches have often viewed time simply as a tool for measurement (reaction time or response time). The longer the time taken to respond to a stimulus, the more it has seemed natural to consider that the situation was complex to resolve or that the intellectual process being studied was deficient (this may be perception, reasoning or any other activity).

More recently, time has become an object of study in its own right once again rather than simply a tool for measurement. The work done on dynamic environments has catalysed this revised approach [48].

Time is a safety management tool from two perspectives:

- on the one hand it is encoded in the representation of the activity itself and serves as a temporal indicator for the organisation of work. De Keyser [22] introduces the concept of temporal reference systems to demonstrate the existence of completely different time-scales, which evolve in parallel when doing professional work: some on a scale of seconds and others on a scale of months. The operator often makes contact with his limits, using these as reference values against which to organise his activity and divide it up over time. These multiple limits may sometimes lead him astray, but in the vast majority of cases the operator manages these parallel-time systems very well and uses them as natural markers to divide up his work during the day;
- on the other hand, time is a driver of transformation in the world and has its own potential to resolve problems and errors. Since situations are dynamic, the key problem at a given moment is usually not the same as the key problem later on; many difficulties can thus be resolved by doing nothing. In the same way, as time alters the situation and automatically leads to the accumulation of information, in many cases it transforms a complex problem into a simple problem, particularly in highly instrumented systems; human beings know this and constantly make use of this property. It is easier to manage prototypical situations, in which the reflex responses are well-known and effective, than to manage situations in flux, where it is necessary first to invest in understanding and where there is a higher risk of taking the wrong action.

Air traffic controllers leave "time to time" to simplify their work. Morineau [49] shows, for example, when studying situations in air traffic control, that controllers only trigger the conflict processing system once all the elements of the conflict are present on the screen and all the means of possible action are available to them. In many cases the conflict has already been seen for some time but it is difficult to characterise under such partial conditions and it is often impossible to correct it using the simplest methods if the controller takes precipitate action as soon as he sees it; waiting presents obvious advantages, including in terms of managing the workload. Finally, the structure of the information system (the control screens)[7] has been designed to allow the operator to have a comfortable margin within which to manage conflicts.

Precisely the same thing applies to the control of errors; time is often a valuable tool when it comes to cataloguing errors and even alleviating their consequences. This property of time is developed during the next paragraph using a number of examples, since it forms the basis for the ecological regulation of risks.

Controlling the Time Required for Understanding

If operators have a choice, they prefer action to understanding, because action aids understanding. These ways of making things simple and spontaneous often clash with the safety divisions within industries and form a focus of ongoing conflict. This is the case in the recent development of the concept of situation awareness, which, when it is understood and used incorrectly by those in charge of safety procedures, suggests that the situation should be completely understood at all times before taking action [50]. This is simply impossible for the operator, and it is even dangerous in terms of managing the process, since the speed at which an exhaustive representation of the world can be built up is much slower than the speed at which the situation changes, and the result will in many cases be a degree of slowness in implementation that would produce an ideal situation but would do so outside the time available for intervention. Only very slow-moving processes such as those in the nuclear industry (and even then only in certain cases) can accommodate an instruction to "take no action" for a predetermined period of time which is reserved for reflection in the immediate period after an incident occurs. A number of experiments in relation to more fast-moving processes, such as in aviation [51], show that pilots exposed to failures do not seek to obtain complete understanding before taking action; on the contrary they limit their analysis and prefer to take action in the direction of the goal that is still compatible with the changing nature of the situation.

[7] In the example of air traffic control, the radar screen provides a zoom view of the situation in which aircraft take several minutes to cross the screen from one side to the other; the screen also has distance reference markers shown as concentric circles, making it easier to position the aircraft in relation to each other.

Controlling the Time Taken for Error Management

Confidence [47] in the ability to control risks lies at the heart of the cognitive control of risk.

In a thesis on the regulation of calls to the “Samu 75” emergency ambulance, Marc and Amalbert [52] studied the contribution made by each member of the team towards the group’s safety. Particular attention was paid to the time that had elapsed and the criteria for intervention by individuals (reporting to others, recovering from a personal decision) required to recover from a risk detected in the collective situation. The results showed that telephone operators often wait several minutes between the time when they perceive that slips are accumulating in the handling of calls and the time when they intervene to correct those slips; they tend to intervene through successive “nudges”, raising the alert level within the group before actually initiating a high alert or attempting a recovery. Everything seems to be done as if the operators accept that the group is constantly drifting towards a significant level of risk and intervene at the limits of that level of risk to keep it manageable and reversible. A number of rationalisations can be found for these behaviours: interactive management of their own workload and their activity for the benefit of the group, controlling the number of times others are interrupted to limit other risks, or confidence in time or other actors correcting problems [53, 54].

The use of time to facilitate work in general medicine [53, 54].
A recent study looked at almost 1,000 cases involving complaints in general medicine from the perspective of management of time and the associated risk of errors.

Medical work involves managing a huge variety of cases and situations, which require different types of anticipation to keep the patient and the practice under control. The analysis proposes four different sources of time, or tempos, each of which has to be controlled since it represents its own risks, as well as intervention at the overall level (synchronously in all four tempos) to retain overall control of the situation.

The doctor’s skill consists of playing with time rather than being caught by time. Time reveals the evidence. The longer one waits, the more evolving phenomena will reveal themselves. Playing with time is therefore fundamental, particularly in primary care, where patients have diseases which are more likely to be at an early stage.

Nevertheless, the time saved on one of these dimensions mentioned above is always reused for the benefit of another dimension (it is even possible to speak of time credits). An older person who comes in for a repeat prescription will not be undressed, and the time saved may help to offer a half-hour explanation when breaking difficult news to a young patient afterwards, or social time saved at home can be paid back by going home earlier than usual. All times and all tempos are exchanged dynamically; what is given to one is taken from the others, automatically, while the total passage of time is under the control of external physical laws. This management process can be helped, it is far from intuitive and if it becomes chronically out of control this can lead to an explosion of medical errors.

The four tempos that were identified are:

- **disease and treatment tempo**. This places the patient in a box of time available to guide the actions involving that patient, see him again, transfer him elsewhere or manage him using important external feedback loops. The doctor knows, for example, that most cancers progress quite slowly, so that they can be expored without an admission to hospital, without taking an inappropriate risk that they will become worse over a period of 1–3 months (the time taken to complete the whole workup);
- **the patient's time**. The patient controls part of the agenda over his own illness; he decides to express his symptoms in ways that are very much weighted to suit his own personality and anxieties. Patients often complain about delays in diagnosis but studies have shown that they are very often partly responsible for those delays, are slow to present their needs, and neglect to arrange the prescribed investigations, as their needs are lost amidst a sea of other requirements that are judged to be a higher priority, or simply, by their tone and attitudes, play a part in reducing the level of communication with the doctor [55, 56];
- **the consultation tempo** is most familiar. It comprises (1) the examination time, (2) the constant interruptions, whether from the telephone, additional illnesses, impromptu visits, (3) all the administrative time that has to be fitted in during the day, (4) and private time. Medicine is only one part of life and it is often necessary to give immediate priority to time for private life, fitted into the professional diary;
- **medical system time**. Patients in primary care are free agents and all prescriptions for examinations or specialist consultations are ballistic in nature. One can never completely know when the patient will come back, make an appointment, receive the results etc.

Summary: A Model of Individual Safety Based on Constantly Building Compromises

The foundations of what could be called a theory of ecological safety [21] shed additional light on some of the findings from the literature on the control of dynamic situations.

The key to interpreting results coherently is based on the following points:

- controlling the situation demands supervision on two levels: supervision of both the external process and the mental process;

- the priority of cognitive activities in the conscious domain is to supervise the workload and ensure coherent progress is being made towards the goal; once this has been ensured, supervision of the physical process can take place at a relatively routine level, using highly proceduralised forms of know-how.

In short, when the work is securely under control, supervision of the physical process is largely automated while control of the situation (internal supervision) paradoxically requires a constant brake on oneself to avoid being tempted to pointlessly engage with local optimum standards (perfect understanding or perfect actions) which are disconnected from the social demands and goals of the situation.

Safety within every supervisory process is ensured by a number of different cognitive mechanisms:

- in the case of external supervision (supervising the result achieved), the routines incorporate a first independent level of checking and adjustment. The threshold for triggering these checks arises at a relatively late stage, so it is necessary for a significant drift in the values for the physical process to occur before this (often automatically) activates a correction routine. The more obviously and the more quickly the situation is drifting (while remaining within the usual reasonable limits), the easier it will therefore be to trigger a correction for the routine (example: monitoring the vehicle's path in the lateral and horizontal planes when driving a car). Conversely, the less obviously and the less quickly the drift becomes visible, the more time, resources and deployment of internal supervision will be required to trigger the correction to resolve a problem which is not a standard, routine one;
- internal supervision (watching oneself work) manages the attention-related activities that are needed to coordinate this process. It also has to marshal its resources and make the best possible trade-offs to achieve "adequacy" in a way that is compatible with its resources. Not every doubtful point can be understood in depth, and the time available (before action becomes necessary) does not very often make it possible to explore all the solutions that are known and available. Flirting with an experience of risk that remains controllable becomes a tool for moment-by-moment cognitive management. As in the case of external supervision, but in this case using a different mechanism, the tactical control of cognition is based on the time that is left before the deadline and the turbulent limits of the cognitive system (Author's note: the concept of turbulent limits is taken from Gibson's vocabulary), where these limits are signalled by the emergence of warning signals indicating the imminent loss of control. These alerts reflect the awareness of difficulties with internal supervision: too many errors, too much time taken to detect errors, too much self-censorship to understand given the lack of time and resources (while the subject is certain that just a little time would be sufficient to gain an understanding), in brief: a feeling of quantitative overload in terms of the action that has to be taken. Through experience and learning, these signals will occur long before the actual loss of control, as soon as the first difficulties are sensed (concept of margins). When they occur, a change of strategy occurs and the operator switches to a different mode of control, which most commonly consists of revising his target contract.

The ecological safety model: a cognitive investment that is adequate for the aims being pursued. An understanding of the human brain shows that it works in an extremely reliable and sophisticated way, quite the opposite of the message of inadequacy and unreliability which is generally associated with human behaviour. This is due to a misunderstanding and is the result of studies that are too focused on numerous errors (which are poorly understood) and rare accidents (which are subjected to an excessive amount of study).

We should recall here that if almost 80 % of serious accidents in high-risk industries have a human cause, we also see 99.9999 % of working situations without a serious accident,[8] a result which is mostly achieved thanks to the astonishing cognitive abilities of the operator.

There is an urgent need to draw lessons from this in terms of the expectations that are placed on operators and to carry out an in-depth review of the indicators and methods used in diagnosing safety, in particular reintroducing the study of "normal" situations and avoiding the use of the reductionist prism of errors as a key variable for analysis.

Consequence: Following Procedures Means Being Able to Deviate from them About an Average Point

A naïve vision of "ideal" cognition wants the operator to achieve greater safety and reduce errors, if he can only be induced to follow the predetermined procedure to the letter, with no deviation at all.

This expectation is extremely naïve and it is never met, for at least two reasons:

- a sociological reason, based on the idea of the "Making of Safety" proposed by de Tersac and Mignar [57] in his analysis of the disaster at AZF (a disastrous factory explosion in Toulouse which occurred on 21 September 2001, 10 days after the Twin Towers attack), "the safety rules are based on a process of organisation that cannot be reduced to defining procedures to be complied with and less still to recording divergences or breaches, but which instead involves inventing rules of practice that complement the formal rules—themselves nothing but "paper rules" because those for whom they are intended do not put them into practice" (page 10). The transition from these published rules, which were designed by the *happy few* in the management and the safety division, to the rules that are applied by all, requires the creation of unwritten social rules on forming a consensus for their unwritten acceptance and application and an interpretation by each player in their own context of what is or is not acceptable in

[8] In the majority of major high-risk industries (nuclear industry, transport etc.) the disaster rate is below 1 per 1 million (1×10^6) units of activity measurement (for example airport movements, or passenger-kilometres by rail).

terms of application of the rule and the deviation that is tolerated. The author gives a useful term for this structure: making safety;

- a cognitive reason, linked to the model described in the pages above: even if the process of "making safety" reaches the conclusion that the formal rule should be followed with no divergence, this would simply be impossible for the operator and would quickly become intolerable even for the management. Following the rule without deviating at all would automatically result in a fall in performance due to the requirement for additional checking which it imposes on the cognitive control process (disengagement of routines, return to a more controlled and slower way of working). The resulting slowing of production would be quite considerable, close to the level of performance at beginner level, and would no doubt be inferior to the "normal and expected" output of an expert worker/operator by a factor of two or three! What is more, such an approach with no deviation would correspond to a lack of feedback on the situation for the operator (decoupling from the real situation and loss of sensitivity due to working too far from the "turbulent and informative" limits of the environment), would be particularly inconvenient to use, and would result in a reduction in his vigilance and his natural recovery mechanisms, exposing him to a slow drift in parameters that would not be perceived until it was too late [58, 59].

The cognitive system is not really capable of managing its internal and external risks effectively without coming into contact with them; seeking to forbid the operator to experience these risks and imposing a process on him that allows no deviation is nonsense in both psychological and ergonomic terms.

Of course the benefit of this constant search for exposure to micro-scale variations in the environment in order to control them better only has its effect within an envelope of levels of risks (and errors) that are agreed and habitual (one might say everyday), in which the operator has the know-how to recover from these routinely. We are not talking about exposure to levels of risk that go beyond the competence of the operator or take him completely by surprise.

The Complex Links Between Safety and Competencies: An Inverted U Curve

The representation of one's own competencies (metacognition) is another variable that determines the successful control of a situation. The criteria that the subject imposes on himself in terms of the objective (the initial contract between himself and the enterprise) influences all strategies and tactics used in supervision and provides the first level of control over the degree of risk that will be accepted when doing the work.

This model of the control of risk-taking which is constantly adjusted according to reflections on the actor's competence and confidence in himself cannot fail to benefit safety.

In fact, in such a system, the more technical competency the operator acquires, the more successes he accumulates, validating his mastery of the situation, the more his routines integrate the capacity for recovery and adjust themselves (this is an automatic, irrepressible process and the subject is not aware of it) by seeking out the turbulent limits of the more remote environment (deviation signals) in order to exercise self-control. The expert operator therefore carries out his process in a routine way, without even being aware of it, with more deviations from the rule than the less expert operator.

This is not, however, worse in terms of safety in managing competencies. In conscious processes (since the last point only concerned the control of routine processes), the expert, fully trained operator gradually adjusts his performance contract in accordance with his successes and failures. The more successes he accumulates, the more his cognition integrates the idea that he can increase his performance contract if the situation requires it. This cognitive feedback mechanism is automatic and mostly irrepressible. To some extent, success feeds the representation of the expert's knowledge in return, promotes the increase in his confidence and automatically encourages him to take greater risks and to seek to validate his knowledge "one step further". In return, recognition of his success by the enterprise or by society (hero status, or at the very least expert status) gradually reinforces a certain level of demand which he places on himself in the way that he works in future (to show that he truly does have this expert status).

This mechanism of self-reinforcing confidence lies at the basis of learning and for a long time contributes towards safety along the learning curve (gradual reduction of errors, increasing confidence); but it does not reach an end (expertise is infinite) and above all, it has little to do with external regulatory constraints.

To some extent, regardless of the rules that are imposed on him, the more technical competence in his work an operator acquires, the more his cognition will integrate the fact that he is able to cope with higher risks in order to achieve higher performance in his work; he will do this first in expected circumstances (reach the professional level) and then in circumstances where he is required to do exceptional things for which his expertise will be valued, and then increasingly on a routine basis in circumstances that do not require it, well beyond what a reasonable approach to safety would require. The more society, the enterprise or those in authority over him "encourage and celebrate" him for this level of performance, the more the expert operator will seek to go one step further when the occasion allows it.

The shape of the relationship between safety and competency is therefore an inverted U curve.

Under these conditions, it can be seen that training an expert to be able to act in rare and technically difficult circumstances does not automatically result in an increase in safety; in fact the opposite is true. The expert who is trained in this way will achieve higher performance, but will be accustomed to risk, his cognition will sublimate it and he will use it in everyday situations even when the enterprise does not wish it.

Ultra-safe systems have understood this and have voluntarily moved away from training operators to deliver exceptional performance so that they are not exposed to over-confidence, divergence and excessive risk-taking in normal situations. The aviation industry has decided, for example, not to train its pilots in non-standard manoeuvres, in particular recovering a passenger aircraft from a bank in excess of 45°, or a modern aircraft from a stall, on the basis that training them in these manoeuvres would only be useful in exceptional circumstances (less than once per 10 million flight hours…. A pilot flies 300–500 h per year) but would have the effect of excessively increasing pilots' confidence on all flights and would result in these difficult manoeuvres being carried out even when they are not necessary.

To some extent, in view of the understanding of the risk management characteristics of human cognition, which adjusts itself with no upper limit to what is perceived as being under control, it is necessary to be very clear about the objectives and types of training offered to operators.

- If one wants to train experts to be capable of exceptional performance (special intervention forces, fighter pilots, surgeons and doctors working in departments that are known for their major innovations), providing training and exposure to increasingly difficult situations is the right thing to do. However, in this case the safety dimension will be sacrificed in that the number of undesirable events will be higher than if simply competent operators perform routine procedures.
- If one wishes to train operators who will routinely comply with a performance standards specified by the organisation, it is better to avoid training operators to become "super-experts" who are capable of managing exceptional levels of risk. This approach applies in the majority of professional environments.

Exceptional competence is associated with increased risk-taking. A very interesting study was published in 2004 on profiles of the victims of avalanches occurring from 1972 to 2002 in the United States [60]. More than 75 % of avalanches resulting in fatalities occur in very high risk locations and conditions and are largely predictable, known and announced by all local media on the relevant stations. Almost 70 % of the groups who died had one or more experts among them who were: accustomed to difficult winter conditions (24 %), expert amateurs trained in avalanche survival (28 %) or even high mountain guides teaching avalanche survival (15 %). This proportion of (very) high competency levels in the groups of victims is much higher than the standards for the groups who regularly engaged in off-piste and high mountain activities during the same period (1972–2002). The groups that suffered accidents had larger numbers on average (8–10) than the exposed groups that did not have victims (2–4); more of them had a well-known, charismatic leader, they were known for having frequently succeeded in overcoming the same difficulties or equivalent difficulties in the past, and on the day of the fatal avalanche more of them had found themselves in situations where difficult decisions had to be made requiring great expertise (night falling, changing weather conditions, individual members fatigued or exhausted, calling into question a route that was safe but much longer etc.). In brief, this study perfectly demonstrates the mechanism described in this paragraph: top experts increase their risk-taking behaviour, which has long been valued because of the exceptional performance that results from it, before they are punished by disastrous accidents. In the area of mountaineering, the story is simple. “For example, it is necessary to use figures to demonstrate the level of madness that is induced by K2: between 24 June and 4 August 1986, 27 mountaineers, all exceptional experts in their field and many of them with global reputations, reached the summit of K2. Thirteen individuals died, 10 of them after successfully reaching the summit or coming very close to doing so. During the next five years, five more “summiters” died on the mountain. Wanda Rutkiewitz was the first woman to successfully make the ascent. She died in 1992 on Kangchenjunga. Five women have climbed K2, and all of them have died in the mountains. This has continued to be repeated. In 1995, six mountaineers including Allison Hargreaves from the United Kingdom, who had just become the first woman to successfully climb Everest without oxygen, were caught in a severe storm above the bottleneck during the descent from the summit of K2. They all died. “Charlie Buffet. Le Monde, 30 August 2001

In the final analysis, the deployment of individual skill to control risk can be expressed by a few practical paradoxes, which were put into words long ago by Dörner [61].

- The sense that supervision is well under control is expressed in moment-to-moment performance which is often imperfect, but where the operator knows that he can achieve an ambitious target using his own personal know-how or the collective know-how of others whom he can rely on. Anticipation “looks from a distance”, the error flow is quite large (mostly routine errors) and the understanding of the situation is limited to what is strictly necessary, freeing up resources for other tasks and in particular for strategic guidance towards the goal; immediate tactical guidance is entrusted to routines that are linked to the environment. Cognitive “copying[9]“ (the result achieved at each moment during the work) can therefore be understood as an exercise that is still incomplete when the subject is fully in command of his situation. The subject is aware that he has not (yet) done everything that should have been done, and that he has made mistakes here and there that he has not yet recovered. This sphere of awareness of “incompleteness” orders his cognitive priorities and often accounts

[9] The term “copying” is used metaphorically here to express the idea of a student's copy that is submitted to his master.

for the moment-by-moment deviations which have the sole purpose of gaining more time in order to recover from delays. This concept of an incomplete draft is indispensable to the dynamic management of cognition and proves effective in terms of reaching the target (despite all these imperfections at each moment, the final result is usually correct); but it also creates many difficulties when it comes to designing and engaging with aids, since these are frequently very directive in terms of correcting faults immediately and despite trying to support the control of this process of dynamic risk management, they seriously disturb it.

- Paradoxically, when such an operator begins to doubt his sense of being in control (which does not mean that the operator has already lost control) he has a feeling of cognitive overload which is expressed in a reduction in his "behavioural waste": he "pays more attention", makes fewer errors, returns to the standard handling of the solution that he thought was effective, anticipates "less far ahead", slows down his work, reduces his ambition, reduces parallel working (in particular thoughts from his personal life) and engages in intense activity aimed at looking for alternative solutions (giving priority to linear inferences, which are often ineffective). It should also be noted, quite paradoxically: this behaviour is often judged to be more reassuring and safer by external audits than the in-control behaviour described previously, as long as the operator follows procedures and adheres to instructions more closely. The operator is frequently aware of this expectation and adopts this behaviour whenever he knows that he is being observed or assessed.
- When he has completely lost control, the operator turns his attention to a part of the problem which is fully under his control and in which he does not make any errors (searching for reassurance) but the rest of the situation and the final outcome of the problem is abandoned (perhaps entrusted to the group or to an automated system instead).

What Lessons Can Be Drawn From This?

The ecological, individual and spontaneous safety model and risk management approach that emerges from this work does not guarantee total safety. It carries within it the seeds of errors that can potentially be very serious. It does, however, make it possible to understand these errors in a different way from traditional error models.

The underlying hypothesis is based on a cognitive system that "wants to survive" and equips itself with the resources to ensure its own safety. It does, however, also need to be effective; a maximalist position with complete, constant control over performance considerably reduces potential performance. The cognitive system is configured dynamically in order to respond to these two contradictory objectives. This configuration is based on two pillars: (1) relying on routines and their automatic linkages to the environment in order to make tactical corrections when cognition reaches the initial limits of controllability (which are still easy to recover, and thus still allow some margin); (2) relying on metacognition (a perspective on of his own competencies) to manage the strategic aspect and keep the target contract within an achievable area (through experience).

Severe errors may occur when these cognitive pillars are hollowed out, either because the signals from the environment are masked or because metacognition indicates an ability to manage the situation which is incorrect and too ambitious. These two conditions have often been met at the beginning of the process of automating systems: automatic processes masked the loss of cognitive control by guaranteeing maximum performance even without any intervention or understanding on the part of the operator; the operator's knowledge of the system became more heterogeneous due to the increase in overall complexity. The mechanisms of memory and meta-knowledge ultimately eliminate part of this heterogeneity and allow the operator to believe that he knows more than the reality of his cognition [62].

Hollnagel [63] used these same ideas in his ETTO (efficiency thoroughness trade-off) model. This strongly sets out the benefits, including safety benefits, that can be achieved by relying on the spontaneous functioning of the operator, which is effective and anticipates to a considerable extent but relies heavily on routines and is exposed to errors (most of which are recovered). This is preferable to asking him constantly to work contrary to his natural disposition, adopting an excessively meticulous approach, imposing procedures and diverting his attention to the very short term, slowing him down and ultimately making him commit more severe errors due to neglect of the medium and long term. For Hollnagel, progress in safety involves studying and optimising these natural human capacities, which have been clearly shown to exist in normal situations and make it possible to achieve a remarkable level of safety as expressed by a very high level of avoidance and recovery from situations that do arise (the positive side that people do not see) rather than studying errors and faults (the negative side… which is ultimately very low-volume and inexorably destined to become even more marginal as progress continues, is difficult to study and is subject to analysis bias).

What are the Consequences of Improving Safety on this Individual Scale?

- It is necessary to avoid the misuse of language when defining errors; incidents and divergences are only measured (and seen) if they are judged to be culpable and the mistake is made of equating this frequency of incidents with the frequency of errors. That is false. There are 100–1,000 times more errors than the number of incidents seen and recorded in a factory or a hospital… but the vast majority of them have been recovered before causing a recordable incident. This misuse of language ends up having a negative impact on safety: it is unrealistic, it cannot be heard by the operator and it is unfair because it minimises the recovery from (near) incidents.
- Safety does not consist in eliminating all errors (that would be a Utopian aim), but in reducing the number of incidents and accidents and errors that have an impact on the process.

- Safety does not consist in adhering to an ideal, imposed process that leaves no flexibility to the operator on either side of the recommended action. Creating a safe working situation (1) first of all means designing a working situation that maximises cognitive "value", reduces the cognitive burden on the operator, allows him to work at his best by using his natural capacity to control risk, anticipate, and thus allowing him to express his ability to recover and use his intelligence throughout the time from the beginning of his shift to the end and (2) also means designing a situation that allows sufficient production to take place, compatible with the economic and social imperatives (what would be the use or benefit in terms of risk of designing a way of driving a car which is absolutely safe but can never exceed 50 km/hr; the macroeconomic and productivity losses would be much greater than the local benefit). It is therefore necessary to be able to permit an error rate consistent with this process of offsetting risks by concentrating the safety process on their recovery.
- Competence promotes safety up to a certain point (inverted U curve). Continuing training at rising levels of risk beyond the risks encountered under usual working conditions (which includes the range of common poor conditions) makes it possible to train "super-experts", but in turn creates the risk of a deterioration in safety due to excessive risk-taking.
- In this necessary system of compromises, attention must be paid to those aspects that could severely destabilise the control and use of routines by professionals. Particular attention must be paid to situations where operators in temporary placements are exposed to unfamiliar situations. These situations require special vigilance in terms of workplace design. We will discuss this again when we look at the more integrated perspective on the approach to the workplace in the next chapter.

References

1. Lewin K (1935) A dynamic theory of personality. McGraw-Hill, New York
2. Duncker K (1945) On problem solving. Psychol Monogr 58:270
3. Broadbent D (1958) Perception and communication. Pergamon Press, London
4. Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97
5. Shiffrin R, Schneider W (1977) Controlled and automatic human information processing: perceptual learning, automatic attending and a general theory. Psychol Rev 84:127–190
6. Wickens C (1984) Varieties of attention. Academic Press, New York
7. Norman DA, Shallice T (1986) Attention to action: willed and automatic control of behavior. In: Davidson GS, Shapiro D (eds) Consciousness and self-regulation, vol 4. New York, Plenum Press, pp 1–18
8. Norman D (1981) Categorization of action slips. Psychol Rev 88:1–15
9. Rasmussen J (1983) Skills, rules, knowledge: signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans Syst Man Cybern 13:257–266
10. Reason J (1990) Human error. Cambridge University Press, Cambridge. (trans: French by PUF, Paris, 1993); (trans: Spanish, Modus Laborandi, 2009)
11. Fioratou E, Flin R, Glavin R (2012) No simple fix for fixation errors: cognitive processes and their clinical implications. Anesthesia 65:61–69

12. Besnard D, Greathead D, Baxter G (2004) When mental models go wrong: co-occurrences in dynamic, critical systems. Int J Hum Comput Stud 60:117–128
13. Hayes J, Flower L (1980) Identifying the organization of writing processes. In: Gregg L, Steinberg E (eds) Cognitive processes in writing, Lawrence Erlbaum Associates, Hillsdale
14. Allwood C, Montgomery H (1982) Detection errors in statistical problem solving. Scand J Psychol 23:131–143
15. Allwood CM (1984) Error detection processes in statistical problem solving. Cognitive Sci 8:413–437
16. Wioland L, Amalberti R (1996) When errors serve safety: towards a model of ecological safety. In: Hollnagel E (ed) First Asian conference on cognitive systems engineering in process control (CSEP 96), Japan, pp 184–191
17. Doireau P, Wioland L, Amalberti R (1997) La détection des erreurs par des opérateurs extérieurs à l'action: le cas du pilotage d'avion. Le Travail Humain 60:131–153
18. Rizzo A, Bagnara S, Visciola M (1987) Human error detection process. Int J Man Mach Stud 27:555–570
19. Amalberti R, Wioland L (1997) Human error in aviation. Keynote address at the International Aviation Safety Conference 1997 (IASC-97), Rotterdam Airport. In: Shoekkha H (ed) Aviation Safety, VSP BV: The Netherlands, pp 91–108
20. Helmreich R (2000) On error management: lessons from aviation. Br Med J 320:781–785
21. Amalberti R (2001) La conduite des systèmes à risques. PUF, Paris
22. De Keyser V (1996) Les erreurs temporelles et les aides techniques. In: Cellier JM, De Keyser V, Valot C (eds) La gestion du temps dans les environnements dynamiques, PUF, Paris, pp 287–310
23. Ghaferi A, Birkmeyer J, Dimick J (2009) Variation in hospital mortality associated with inpatient surgery. N Engl J Med 361:1368–1375
24. Dekker S (2006) Ten questions about human errors. Avebury-Ashgate Publisher, Hants
25. Johnson C, Holloway C (2004) Systemic failures and human error in Canadian aviation reports between 1996 and 2002. In: Pritchett A, Jackson A (eds) HCI in aerospace 2004, Eurisco, Toulouse, pp 25–32
26. Johnson C (2004) Human error and the failure of imagination, In: Johnson CW, Palanque P (eds) Human error, safety and systems development, Preface, Kluwer Academic Press, New York
27. Norman D (1983) Some observations on mental models. In: Stevens G, Gentner S (eds) Mental models. LEA, Hillsdale
28. Ochanine D (1981) L'image opérative, actes d'un séminaire et recueil d'articles. Université Paris V
29. Piaget J (1974) La prise de conscience. PUF, Paris
30. Simon H (1982) Models of bounded rationality, vol. 1. MIT Press, Cambridge
31. Falzon P, Amalberti R, Carbonell N (1986) Dialogue control strategies in oral communication. In: Hopper D, Newman IA (eds) The future of command languages: foundations for human computer communication, Elsevier Science Publisher, North Holland, pp 73–98
32. Hoc JM (1988) Cognitive psychology of planning. Academic Press, London
33. O'Hara K, Payne S (1998) The effects of operator implementation cost on planfulness of problem solving and learning. Cogn Psychol 35:34–70
34. O'Hara K, Payne S (1999) Planning and the user interface: the effect of lockout time and error recovery cost. Int J Hum Comput Stud 50:41–59
35. Amalberti R (2001) La maîtrise des situations dynamiques. Psychologie Française 46–2:105–117
36. Amalberti R (2002) Use and misuse of safety models in design. Lect Notes Comput Sci 2485:1–21
37. Klein G, Zsambok CE (1997) Naturalistic decision making. LEA, Mahwah
38. Gibson J (1979) The ecological approach to visual perception. Houghton-Mifflin, Boston
39. Kahneman D, Slovic P, Tversky A (1982) Judgement under uncertainty: heuristics and biases. Cambridge University Press, Cambridge

40. Gibson J, Crooks L (1938) A theoretical field analysis of automobile-driving. Am J Psychol 51:453–471
41. Norman D (1988) The design of everyday things. Double Day Currency, New York
42. Newell A, Simon H (1972) Human problem solving. Prentice Hall, Englewoods Cliffs
43. Maturana H, Varela F (1992) The tree of knowledge, the biological roots of natural understanding. Shambala publications, Boston
44. Zangh J, Norman DA (1994) Representation in distributed cognitive tasks. Cognitive Sci 18:87–122
45. Noizet A, Amalberti R (2000) Le contrôle cognitif des activités routinières des agents de terrain en centrale nucléaire: un double système de contrôle. Revue d'Intell Artificielle 14(1–2):73–92
46. Hoc JM, Amalberti R (2007) Cognitive control dynamics for reaching a satisfying performance in complex dynamic situations. J Cognitive Eng Decis Mak 1:22–55
47. Valot C, Amalberti R (1992) Metaknowledge for time and reliability. Reliab Eng Syst Saf 36:199–206
48. Cellier JM, De Keyser V, Valot C (1996) La gestion du temps dans les environnements dynamiques. PUF, Paris
49. Morineau T, Hoc JM, Denecker P (2003) Cognitive control levels in air traffic radar controller activity. Int J Aviat Psychol 13:107–130
50. Endsley M (1995) Toward a theory of situation awareness in dynamic systems. Hum Factors 37:32–64
51. Plat M, Amalberti R (2000) Experimental crew training to deal with automation surprises. In: Amalberti NSR (ed) Cognitive engineering in the aviation domain, Lawrence Erlbaum Associates, New Jersey, pp 287–308
52. Marc J, Amalberti R (2002) Contribution de l'individu au fonctionnement sûr du collectif: l'exemple de la régulation du SAMU. Le Travail Humain 64:201–220
53. Amalberti R, Brami J (2011) Tempos management in primary care: a key factor for classifying adverse events, and improving quality and safety. BMJ quality and safety online first, published on 2 Sept 2011 as doi: 10.1136/bmjqs.2010.048710
54. Brami J, Amalberti R (2009) Les risques en médecine générale. Springer, France
55. Barber N (2002) Should we consider non-compliance a medical error? Qual Saf Health Care 11:81–84
56. Buetow S, Kiata L, Liew T et al (2009) Patient error: a preliminary taxonomy. Ann Fam Med 7:223–231
57. De Tersac G, Mignard J (2011) Les paradoxes de la sécurité, le cas d'AZF. PUF, Paris
58. Rasmussen J (1997) Risk management in a dynamic society. Saf Sci 27:183–214
59. Polet P, Vanderhaegen F, Amalberti R (2003) Modelling the border-line tolerated conditions of use. Saf Sci 41:111–136
60. McCammmon I (2004) Heuristics traps in recreational avalanche accidents: evidence and implications, vol 68. Avalanche News
61. Dorner D (1997) The logic of failure: recognizing and avoiding error in complex situations. Perseus Books, New York
62. Amalberti R (1998) Automation in aviation: a human factors perspective. In: de JWDH D (ed) Aviation human factors, Lawrence Erlbaum Associates, Hillsdale-New Jersey, pp 173–192
63. Hollnagel E (2009) The ETTO principle. Efficiency-thoroughness trade-off. Ashgate Publishing, Farnham

Chapter 3
The Keys to a Successful Systemic Approach to Risk Management

Abstract It is universally admitted that an approach to safety applied to our complex industries (nuclear, chemical, construction and skilled trades) and services (medicine, banking and finance, public and private transport), can no longer be limited to finding local technical solutions; it absolutely must be systemic and global. How should these concepts be fleshed out? This chapter seeks to answer this question from various different perspectives, using examples taken from many contrasting areas, breaking down bias and prejudice and offering practical keys.

On Safety, Systems, Complexity … and the Structure of this Chapter

The management of risk in an enterprise is not only about avoiding or reducing accidents (affecting the system or those who work in it). It also concerns everything that may compromise survival, whether the threat is economic, political, social or damage to the image of the enterprise, particularly following an accident.

In order to understand a systematic approach, one must accept that risk management covers all the risks that could "kill" the enterprise, whether they are social, technical or financial.

The reduction of risks in a socio-professional system is therefore a complex concept, which can be defined differently depending on which perspective is adopted: fewer industrial accidents, fewer accidents affecting the installation, fewer risks of harm to social conditions and operations (no redundancies, protecting careers) or fewer risks to the business model (debt, profits, economic vulnerabilities).

Risk management covers all the risks that could "kill" the enterprise. Safety, in terms of avoiding accidents, is only one such risk: other (economic and strategic) risks can sometimes kill the enterprise more quickly and may therefore take priority over safety when it comes to short-term investment.

In this case, the art of risk management is to set priorities, make trade-offs, manage emergencies (doing effectively what one has decided to do),

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology, DOI: 10.1007/978-94-007-6549-8_3, © The Author(s) 2013

but also remembering that some areas are neglected as a result and managing these in a specific way within the relevant divisions (a sound awareness of what is not being done, through the development of an awareness of these temporary vulnerabilities among both managers and operators, for example by strengthening detection and recovery systems when it is not possible to make further investments in prevention).

All these dimensions are legitimate, even though they often conflict with each other: the economic survival of the *business model* often involves increased exposure to the risk of an accident, which is managed more or less rationally and effectively after a setback (Fukushima is the extreme example of this).

This text endorses the perspectives of industrial safety and safe service provision (reduction of accidents, patient safety in medicine) by showing how the literature and experience on the ground are now making it possible to build a systemic approach which is effective, coherent and maintains control of the compromises that are made in relation to other areas of risk within the enterprise.

The key to the success of the systemic approach can be summarised by three complementary key points: (1) controlling the four stages of the trade-offs which are always present in building the safety structure of a complex system, (2) doing well what one has decided to do, and knowing and controlling what one has decided not to do, (3) future thinking rather than past thinking.

This text is structured around those three key points

The Swiss Cheese Model as the Archetype for Systemic Models … and Its Current Limitations

When speaking about the systemic model of risk management, everyone immediately thinks of the slices model set out by Reason in the 1980s [1, 2].

This model, based on three tenses, is simple and tells us: (1) that one cannot completely eliminate (patent) errors by people who are directly engaged in work, (2) that deep defences are needed to avoid the propagation of these errors as far as an accident, and (3) that it is necessary to be aware of organisational and management errors (latent errors) which, without being the immediate cause of accidents, increase the vulnerability of the individuals and defences directly engaged in the work by not giving them all the resources they need to be effective.

This model is always a heuristic one, and its author, whom I count as both a teacher and a friend, truly deserves his reputation and his global place in the pantheon of those who have contributed towards safety. It must also be admitted, however, that this model is not now sufficient to create a systemic approach that can offer effective safety in complex professional activities. It has four major defects:

- it reflects a linear model of the accident which is based on the propagation of failures in the structures and components making up the model; in this sense it harks back to ideas that are already very old and are inspired by the domino model of Heinrich [3] or to the chain of errors, although it is more complex because it introduces the role of organisations and design[1];

- the model is still profoundly Cartesian, since it breaks down the universe of professional work into parts (structures and components, slices) and then attempts to find the vulnerabilities in each one of these parts; it explains the accident in terms of local vulnerabilities and leads to a search for "the error". The model certainly offers a vision of the whole by referring to the interactions and distortions in the alignment of the slices and vulnerabilities (it is important that failures should not be aligned), but it does not effectively take into account the risks of accidents where there are no failings in the individual parts but which are instead associated with weak links between parts that are not defective and emerging properties and risks of the "whole" (typically the global perspective on the system)[2];
- the model suggests that the identification and complete elimination of latent causes and exposures to risk are the (only) way to make progress in terms of safety. By doing this, it points us towards a model of avoiding or reducing exposure to risk in order to improve safety and eliminate all vulnerabilities, while living with voluntary exposure to risk (voluntarily creating holes in the slices) is a realistic factor that promotes survival for many enterprises. The world of safety must now accept this analysis and not reject it, since sociotechnical systems most commonly die because of their poor economic, organisational and political choices. Reason's model therefore provides keys to action which are valuable and are centred on a simple choice in favour of safety, but which are still inadequate in the real industrial world;
- finally, to reiterate the points above, he remains within the lines of traditional ideas and implicitly supports the idea that the best thing (for the enterprise) is always to achieve more safety, up to the total or virtually total elimination of accidents and incidents. This vision, which is acceptable and makes good sense in less safe systems, paradoxically reaches its limit in systems that have become very safe. The safer a system becomes, the more difficulty it has surviving these last accidents, and the more these last accidents tend to be exceptional accidents which are largely provoked by the system itself, which has become too safe, too rigid, too proceduralised and has, in short, lost its native resilience. We see this every day: ultra-safe industries have much greater difficulty justifying their safety policy (which everyone acknowledges to be effective, despite the very rare accidents that still occur), while fishing and road traffic have difficulty managing safety associated with their continuing succession of daily accidents [6]. Worse still, the search for less and less (accidents, incidents and errors) inspired by the nuclear industry, the aviation industry and the (small number of) ultra-safe industries, ultimately leads us to forget that there are other authentic models of safety (for example the HR0 model or resilience model) which have their own rules and their own contexts for progress, which are very well suited to the thousands of professional activities that are not subject to the demands imposed by our ultra-safe systems.

[1] This criticism is debated particularly effectively in Hollnagel et al. [4].

[2] This criticism is discussed particularly well in Dekker [5].

On this last point, it is also important to understand our own bias when it comes to building our knowledge base in the area of safety, which is associated with the professional applications that are used and studied in order to build these models of safety. Over 20 years (1990–2010) I have reviewed more than 2,072 conceptual articles addressing models of safety and their characteristics, published in eight international journals specialising in the areas of industrial safety, safety at work and safety in services (particularly medical, banking and road safety).[3] More than two-thirds of these articles (1,547) repeatedly address safety (safety at work or process safety) in five major industrial domains (nuclear industry, aviation industry, chemical industry, offshore and construction); almost a quarter of them (483), although in many cases these simply repeat and validate the models from the major industries, refer to medicine, banking and road safety (note from the author: the work done in this last field is rather more original). Very few (42) are original monographs on safety applied to skilled trade activities with low or very low levels of safety (mining, professional fishing, miscellaneous skilled trades). To some extent our understanding of risk and our models of safety largely result from the model of major industry, which is certainly responsible for the disasters highlighted in the media but not for the largest number of deaths, and which is ultimately very homogeneous in terms of its limitations and preoccupations (a level of safety between 10-5 and 10-7, highly developed regulatory requirements, priority given to human and organisational factors etc.); this model taken from major industries ultimately represents only a very small proportion of human professional activity on the planet. This recruitment bias accounts for many of our errors when we make generalisations about ideas relating to the safety of complex systems, by limiting our perspective to a narrow field and almost to a single specific case.

Controlling Systemic Safety: Four Key Steps for Building Safety in a Complex System

Improving safety in a complex system by adopting a systemic approach always requires a procedure that follows the same four steps: (1) knowing where the risks are, prioritising and building an ad-hoc system of defences, (2) setting this paper model alongside the real situation and making adjustments accordingly, particularly to various shifts in practices, (3) carrying out the analysis at a higher level and considering the macroeconomic and political constraints, (4) once all the preceding steps have been completed and the system is much more restricted and constrained, it is still necessary to ask how much resistance it still has to exceptional circumstances; so the question of resilience becomes central.

[3] Search of Google and (review summaries) in December 2011, limited to eight journals: Human factors, Safety Science, Ergonomics, Accident analysis and Prevention, Journal of Safety research, Journal of Risk research, International Journal Quality in Health Care, British Medical Journal Quality and Safety.

The ultimate level of safety observed at time T0 in a socio-technical system is always the result of a four-step construction process. These four steps are successive, but they do have feedback loops.

- The first step is always to identify the risks and establish an ideal model of defence. This is the classical field of risk mapping, decision-making matrices and more broadly systems reliability, extended to include human reliability. Once the risk has been identified and prioritised, this step leads to the definition of lines of defence (barriers) to reduce the occurrence of the accidents that are of concern.
- The second step is to set this ideal model alongside the real situation. In many circumstances, operators do not comply with this model and suffer no particular penalty, at least for a long time; many divergences occur for many different reasons, and it is useful to understand these. This migration in practice will sooner or later lead to incidents and accidents. The end of this phase therefore necessary involves a feedback loop to adjust the ideal model, but the way in which the ideal model is interrogated and corrected in a feedback loop is far from always being identical. In the majority of cases those concerned make the mistake of considering that the ideal model should not be called into question and that it is only necessary to strengthen its defences or the authority exerted over the operator to induce him to follow the instructions and procedures. The idea of strengthening the operator's "safety culture" so that he will follow the script set out in the ideal model that one believes in, is one path which is often adopted in parallel with a purely procedural and regulatory hardening of the ideal model to force compliance with it. Perhaps, however, some of the fundamentals of this ideal model should be questioned rather than seeking to make it stick to the real situation.
- The third step is systemic in nature. No-one imagines that a complex system can be made completely safe simply by relying on putting procedures and recommendations in place and forcing only front-line operators to comply strictly with these good practices and recommendations. A further step needs to be taken to strengthen what one might call the "system", and this step is based on a strategy of "safe governance" of this system, and action involving the middle and top management: how to conceive a safe structure for the system, the relationships between bodies, professions, the specific interests of each one, directorates, divisions, branches and subcontractors, the levels at which each actor should be independent or perhaps dependent, what risks should be taken and how, when controlling the compromises that are made between economy, profitability, certainty and safety.
- The fourth step concerns the ultimate resilience of the resulting model. Safety is often a problem that is even more difficult to control once all the preceding steps have been taken, once the system has become safe or ultra-safe, once the procedures, safety instructions and protective measures have reasonably been followed and adopted by front-line operators, once the management is personally involved in the decisions and trade-offs in favour of safety, sometimes accepting sacrifices in terms of profitability and once a culture of total reporting is established. In fact no safe system can be completely protected from disasters, and although these are

certainly rarer, they are infinitely more damaging to the image of an enterprise or activity that has become safe, to the point that it more easily results in a crisis and sometimes in the death of the enterprise. If the same accident had occurred earlier in the history of the same enterprise, at a time when it was less safe, it no doubt would not have resulted in the same consequences. In the end the excellence of the level of safety that has been reached becomes the mirror for the model which is predicted and looked for; but the system will die of something that has not been predicted. Adaptation to exceptional circumstances is never written into the strictly procedural models. The system is no longer robust in the face of rare challenges and loses its "resilience". This section explains everything that is paradoxical about this final step. The more proceduralisation occurs, the more conformity and safety is achieved in relation to an ideal model and, unfortunately, the more professionals and managers become "deskilled". They are rarely exposed to difficult situations; they lose the habit of making decisions that involve sacrifices in contradictory dimensions (which are characteristic of short-term survival under these difficult conditions). In brief, resilience is a property which is relatively native to less safe systems in which the operators are exposed to highly variable situations. It reduces once the system is made safe using the previous three steps, and at the end of the process it diminishes to the point where it needs to be reinforced using specific mechanisms in systems that have become ultra-safe. Unfortunately this phase of re-inserting resilience is often delicate or even impossible to achieve because it runs counter to the solutions that have been used to strengthen the ideal model and achieve the current level of safety.

We will now look in detail at the content of each one of these steps, what can be achieved by them and the respective challenges facing each one.

Step 1: Evaluate the Risk and Build a Defensive Fortress

It is impossible to initiate a safety process without evaluating the risk and protecting oneself against what appears to be a threat.

The tools used to evaluate risk, both a priori (on the basis of a systematic analysis of the vulnerabilities of the system in question) and a posteriori (based on the analysis of accidents and incidents that have really occurred) are the toolkit of reliability engineers; they are well-known and they are not described in this work. We refer the reader back to the plethora of literature that exists on this subject and we will only cite a small selection of the references, which is no doubt incomplete but is nevertheless sufficient to obtain an overview of the principal methods used.[4]

[4] Numerous references exist on this subject, most of them dating back some time. Many different summaries exist on a number of websites (although this list is not exclusive) http://pachome1.pacific.net.sg/~thk/risk.html accessed on 27 décembre 2011 http://www.statcart.com/ viewed on 27 December 2011or the remarkable summary with reference to the medical field, which was published in French in five articles by a group of authors, specifically [7, 8].

Principal methods of risk analysis

A priori analysis

Analysis of processes	Tool	Usefulness	Limitations
	Process analysis	A prerequisite for other processes (FMEA/FMECA, PRA, HACCP) in order to describe an activity that brings together all the functional constraints (flows, resources etc.) in order to identify critical points and improve the steps in its functions, particularly with regard to interfaces between services	Requires a knowledge of the need that has to be met in relation to the available resources
	Functional analysis	A prerequisite for other processes (FMEA/FMECA, PRA, definition of the need to be met, value analysis) involved in introducing and implementing a new activity (service or product), setting out the functions needed to achieve this, starting from design, applicable constraints and performance criteria	Complex, time-consuming process suitable for designing a new activity forgetting a function
Example of the use of WHAT if type techniques	HAZOP *Hazard and operability study*	Breaking down the system that is to be constructed into parts, referred to as “nodes”, then, using a keyword-based guide, brainstorming on possible deviations, assessment of the potential risks to human safety, property and the environment	This process is well adapted to anticipatory work on new systems. Extensively used in the chemical industry. The quality of the descriptions reflects the quality of the participants
Example of use of WHAT if type techniques	HACCP *Hazard analysis control critical point*	Interactive method involving experts, well codified with 7 steps, applicable to both existing and prospective systems	Authoritative in the food industry
	PRA *Preliminary Risk Analysis*	Identifies accident scenarios in the presence of danger, evaluates them, establishes a hierarchy between them and makes deductions concerning risk control in a relatively exhaustive way	Complex and time-consuming process Applications in ultra-safe or very high-risk industries (space travel)

(continued)

A priori analysis			
Analysis of processes	Tool	Usefulness	Limitations
	FMECA Failure mode effect and criticality analysis FMEA Failure mode and effect analysis	A method to ensure safe functioning which allows methodological analysis of a critical process, from description of the steps involving risk to measurement of the criticality of causes (frequency multiplied by severity) FMEA: a simplified approach to the FMECA method which can be used in the absence of quantified data	Selection of critical processes A lengthy process, reserved for intrinsically very high-risk processes
Instrument panel monitoring	Instrument panel warning indicators	Possible method for monitoring an action to control risks Makes it possible to detect risks, anticipate and make decisions	Data gathering can be burdensome
Surveys in comparison with a frame of reference	Audit	Verifies the implementation and effectiveness of defined measures Simple to understand *Specific example: the LOSA Line-Oriented Safety Audit: which is the method recommended by the ICAO, relies on auditors seated behind the pilots on a jump seat, entering data on errors made by the crew* Other examples: Audits of human and organisational factors	Requires programming and preparatory work (auditor training) and a validated, professional frame of reference. Not well suited for organisational analysis The subsequent methodological analysis is essential in order to benefit from any observation of a dysfunction
	Risk visits, Walk rounds Health And Working Conditions Monitoring Committee	Lever for possible triggering of an institutional process (for example insurance) and action by the management One particularly well-known version of such visits is applicable to action taken in partnership (between trade unions and management) to improve safety in working conditions	Requires follow-up of the observations and a detailed organisation in order to be useful

(continued)

(continued)

A priori analysis

Analysis of processes	Tool	Usefulness	Limitations
Computer-assisted techniques	Go method Markov modelling Dynamic Event logic Analytical methodology	All these techniques use mathematical and computerised formalisation of the process to automatically create graphs showing the propagation of risk starting from an event	Complex and reserved for very specific applications
A posteriori analysis			
Accident analysis or analysis of near misses	In-depth analysis	Looking at the causes of accidents One example which is used particularly frequently in medicine: the ALARM (Association of Litigation and Risk Management) method, which is inspired by Reason's model and focuses on looking for latent errors	Simple but leaves considerable scope for interpretation by analysts
	Tree based techniques: fault tree analysis, event tree analysis, MORT-Management Oversight Risk tree	A series of methods that start from the event and use logical links to establish the causes; the analysis may also start from a sentinel event and seek to understand its potential consequences	Quite time-consuming if the analysis goes beyond the superficial level, requires a favourable environment and sufficient authority to push through the conclusions

In this phase 1 we will only debate a small number of points that give rise to difficulties, which are both specific to this initial phase and strategic in terms of decision-making and trade-offs.

What reference framework should be used to characterise and measure risk?

The analysis, characterisation and measurement of risk call into question our understanding of the scope of the domain that is considered to be relevant for the purpose of explaining risk. To some extent our scientific analysis of risk, which is intended to allow us to measure risk using the most formalised methods possible, depends on subjective values from the outset.

To use a simple (simplistic) example: a bank wishes to produce a map of its financial risks for its "financial products" division. The traditional analysis will cover the range of processes that are used by the trading departments to interact in the markets: organisation of the department, order flows, rules of engagement, relevance of the mathematical risk models used, IT tools, delegation of engagement, supervision and control. If only this domain is covered, one can only have a relatively limited, technical perspective on an internal process within the bank which is specific to trading. It is easy to imagine, however, that the real risk may depend more on global political balances than on trading techniques within the bank. Extending the perimeter of technical risk analysis to include the analysis of political risk at the national level or even globally changes the model of the processes that need to be considered to feed into the charting process, alters the results of the HAZOP, FMECA and PRA models that characterise risk and ultimately changes part of the measurement that is considered to be relevant to this risk.

We understand that an excessively narrow scope of analysis can quite easily result in an analysis of risk that covers only a fraction of the real risk and sometimes quite a marginal fraction.

Taking this one step further: changing to a different risk analysis

Medicine is constantly evaluating the risks and the efficacy of its strategies. Let us consider the strategies for the management of obesity in children. An analysis is carried out of the risk and benefit of the medicines that are given and the management of this problem. It is known, however, that lifestyle habits in underprivileged social contexts and industrial pressure from soft drinks and confectionery manufacturers (which cumulatively promote unbalanced nutrition in schools and at home) represent a potential source of much greater risks in terms of obesity than poor medical management. In this case, the perimeter that needs to be considered in order to prevent the risk of obesity will benefit from consideration of the wider range of risks associated with social action rather than only the risks associated with the limited domain of medical action.

This simple example points to one of the core principles of cost-benefit analysis and economic analyses in the area of safety: how to consider safety from a perspective of greater efficiency, how to achieve better and safer production, at the same cost or if possible even at a lower cost.

This type of analysis is not new, but it is always quite difficult to put into practice because it specifically forces us to extend the perimeter of risk analysis far beyond the limited technical domain that triggers the analysis of risk to its own process in order to evaluate its own work.

Typically this leads to a systemic approach.

For example, an article written some time ago [9] proposes an inventory of 500 high-risk human activities which are evaluated in terms of QUALYs (life-years lost engaging in the activity or passively exposing oneself to risk—those living close to industrial installations or others). Without discussing in detail the method that is used in the comparison, the article does address a fundamental question: everyone carries out their own risk analysis within their own small domain (of benefit) and therefore creates local solutions for risk reduction and risk contingency planning that are sometimes extremely complex and expensive. Taking a "bird's eye view", however, this landscape appears to be divided into hundreds of separate compartments, each one defended from its own risks, which raises questions about the relevance of an approach to the risks that exist within the compartment.

When analysing risk, it should be possible to consider the offsetting of risks and the comparison of risks with other compartments within the same domain, with other alternative solutions (in the case of obesity above… or in a more difficult case with an even more sensitive taboo, the case of risks associated with commercial activity carried out by tele-presence in a virtual environment, as compared with the risk that is accepted in terms of aviation safety, to deal with the same problem by physically transporting operators). Reticence to engage with these thought-processes is understandable, since each model of intra-compartment risks corresponds to a business model that has little interest in a more global perspective that could be harmful to its business. The example of nuclear risks or difficult very deep water oil operations, as compared with alternative energy technologies, shows how it is necessary to go as far as disasters and beyond in order to truly accept this global perspective in a fully transparent way.

The same is true when considering the time-horizon and the capacity for recovery and attenuation. Risk analyses take little account of the often positive trade-offs involved in taking short-term risks in order to safeguard the long term more effectively. Let us imagine a risk matrix which considers an intervention in a factory in a difficult context where there is a leak from a pressurised valve. Acceptance of the immediate risk will affect the long-term risk. Even if the intervention results in an industrial accident in the short term, it may have a considerable cost benefit overall in the long term, since it avoids a more significant breakdown in the plant and no doubt other severe consequences as well. Is it necessary to prohibit taking immediate risks in order to achieve long-term safety in this way? And where is the critical time-horizon?

On closer inspection it becomes clear that managing risks does not always mean reducing them, but it often means exchanging them for other risks and for risks at different times.

These trade-offs result in both gains and losses, depending on the perimeter and the time-horizon under consideration.

For the safety division in the factory, the industrial accident that occurs during a difficult intervention during which safety codes are not complied with, will almost always be considered to be the result of poorly controlled risk management, even if it safeguards the long-term situation. The only exception is in the emotional interpretation of the event, if it is clearly established that it was a heroic action (in other words the benefit is clearly and immediately identifiable in the very short term, for example saving an injured employee in a toxic environment without protection, where the person saving him failed to comply with instructions). Exactly the same logic is applicable to lymphangitis in the arm caused by an infusion of medication which is intended to treat cancer but which one day "bypasses" the vein. The immediate effect is disastrous for the patient: extreme pain, a swollen arm, disability, a number of weeks taken recover, but the long-term effect will be absolutely minor. The overall effect of treating the cancer will be on a much larger scale than this incident along the way which has no longer-term consequences.

Ultimately the difficulty consists in having a system that accepts the dynamic and intelligent nature of such trade-offs. This widening of the perimeter is generally impossible due to the fact that the system primarily consists of human beings, careers, individual attitudes to be justified, conflicting financial interests and power bases, and it is important to recognise that the ultimate beneficiary of the long-term exchange of risks is rarely the person who has to make the decision to accept the short-term risk.

The question of choosing the perimeter can be asked in different forms, including the question of the social judgement that will alter the analysis of the risk acceptance matrix.

Let us imagine an amateur mountaineer, who is walking in the mountains and climbing risky and technically difficult slopes. He exposes himself to a risk which he knows to be very high because his cost-benefit analysis (pleasure) is positive. If this analysis is extended to his family circle, the cost-benefit assessment by his wife will necessarily be different, giving a much lower weighting to the pleasure and putting a higher weighting on the effect of his absence and the risk of an accident. If it is extended to society, however, the cost-benefit analysis will result in a very low positive value (positive economic influence on high mountain resorts), for a very large negative value: the cost of providing assistance, cost of disability. Depending on the chosen perimeter, the analysis therefore leads to different results in terms of acceptance of the risk matrix.

What is the place of voluntary (or compulsory) reporting of incidents during this first phase?

There is a huge amount of literature on difficulties associated with reporting in all industries and services,[5] [10]; the observations are often the same: massive underreporting, due to (1) fear of consequences (sometimes legal, but above all internal

[5] Obstacles to participation: the top 9 reasons why workers don't report near misses, 2011, http://ehstoday.com/safety/news/9-reasons-near-miss-reporting/.

within the enterprise, personal image and sanctions), (2) chronically poor understanding of what has to be reported (a representation of what the managers or the enterprise expects… which filters out many problems because they are judged not to be relevant for these purposes because no-one is affected, they are not severe enough, they have been recovered, they are too common etc.) and (3) ineffective use of the results obtained.

Some solutions have been found to the problem of protection for employees who make reports. A number of legislative frameworks (no blame, no shame) protect staff who make reports in certain countries (particularly the United States and Denmark[6,7]) in many industrial and service sectors. For example, for 25 years the national aviation safety reporting system in the United States (ASRS—Aviation Safety Reporting System) has protected the aviation professionals who report their errors by guaranteeing them anonymity and rendering legal prosecution impossible.[8] Similar systems now exist in the medical domain.

Most of the difficulties associated with the management of voluntary reporting from a safety perspective are different in nature and are still current.

Reporting risk is still very much bound up with the concept of the safety culture and to a lesser extent with the concept of improving results in the area of safety

A system has to agree to be transparent in relation to errors as a prerequisite for becoming safe. Reason [11] mentions four essential features when constructing an effective safety culture, all of which are more or less linked to error reporting: the ability to refrain from punishing those who make reports except in cases of intentional violations with serious consequences (just culture), the ability to share these events that are reported (informed culture), the ability to draw lessons from these reports (learning culture) and the ability to change the organisational model whenever reporting shows the ineffectiveness of the current model (flexible culture). These ideas are now well established and have been dealt with by many other authors[12, 13].

Having said this, reporting and the measurement of the safety culture (which is often firmly centred on this reporting aspect) do raise a fundamental problem in relation to the real link between the amount of reporting and the benefits in terms of the level of safety.

This link is obvious in aviation and in the nuclear industry but more open to debate in other industries.

In fact there is a bias towards specific cases resulting from the model used in civil aviation and in the nuclear industry, characterised by its powerful global, regional and national supervisory bodies (ICAO, EASA, National Aviation Authorities,

[6] Danish act on patient safety, http://www.patientsikkerhed.dk/admin/media/pdf/133907d0940e4d5f751852ec8f6b1795.pdf.

[7] US patient safety and quality improvement act, 2005, http://www.ahrq.gov/qual/psoact.htm.

[8] http://asrs.arc.nasa.gov/overview/immunity.html#, accessed on 26 December 2011.

Nuclear Energy Agency, NISA, IAEA etc.), the reality of total, permanent external surveillance (air traffic control and black boxes). In brief, these are relatively specific systems in which reporting incidents does not leave much of a margin for professionals since they will be seen and read by the supervisors in any case if they cause the slightest consequences. In this sense, the model of civil aviation or the nuclear industry is actually an exemplary model of voluntary reporting. In fact the density of reporting is correlated with the safety of individual airlines in civil aviation, since it expresses a function which is absolutely essential in that environment.

How many other industrial models, however, are similar to that model? Almost none… There is less supervision, actors have greater autonomy; it is not surprising that transverse studies on models of the safety culture whose key aspects are taken from aviation, the nuclear industry and the chemical industry, do not always yield such convincing results in other industries, particularly in medicine.

Reporting and the safety culture: what is the link between the reporting culture and safety performance within the industry? [14–16]

At the organisational level, there are nine dimensions that are repeatedly tested in questionnaires on the safety culture: the first is the policy on risk management and incident reporting, followed by the quality of the technical platform, the quality of maintenance, procedures, the quality and quantity of staff and planning, skills, collective commitment, communication and monitoring change. The use of questionnaires reveals a degree of aggregation and overlapping between the values of these nine dimensions, with the management largely predicting all the other values.

In the end, a number of questions remain unresolved, in particular the formal link between the measurement of a specific type of organisation and the risk of accidents. Values are measured if they are likely to be significant in terms of safety, as in the case of incident reporting, which is largely passed up to the highest level within the organisation, so that there is little independence of scale between the macro, meso and micro levels. One might ask whether, in view of the limitations of these questionnaires, audit techniques would perhaps be more appropriate.

The willingness to submit reports is even more variable in medicine than in the rest of industry [17–22]

All the literature indicates that systems that rely on reporting by health care professionals are subject to massive under-reporting. This is nothing new. In 1995, a study carried out over a 6 month period in a Harvard hospital showed that the reporting rate represented barely 6 % of the actual SAI (Serious Adverse Incident) rate as estimated through retrospective file analysis [18]. In 1998 similar results were found at the Brigham & Women Hospital in Boston: a system that automatically detected SAIs on the basis of electronic patient records had detected 2,620 alerts [19]; after verification, 365 SAIs were identified. On retrospective analysis of the records (which was carried out independently of the previous process) it was possible to identify 385 SAIs while health care professionals had declared 23 SAIs during the same period. Of the 617 separate SAIs detected using at least one of the three methods, 65 % were identified by retrospective analysis, 45 % from electronic records and only 4 % from reports in the official reporting system. The literature consistently reports similar figures (between 3.5 and 10 %), which testifies to the very poor performance of such spontaneous reporting systems. Another result that is often seen [17, 23] is the massive over-representation of reports made by nurses (70–80 % of the databases) as compared to those made by doctors. Among this group, senior doctors were almost entirely missing from the database of staff making reports [21]. The most recent results [22] confirm these difficulties.

Paradoxically, the "no blame no shame" condition and questions about voluntary reporting may become less relevant to the introduction of risk mapping in future

As we have just seen, the question of the lack of legal protection for staff making reports, staff hierarchies and supervisory authorities has been an issue that has been obsessively addressed by the literature.[9] At present, reports from the actors involved represented virtually the only source of information.

Things should be different in future: incident reports contributed by actors on the ground will be marginal in comparison with other means of finding out about deviations and incidents. Due to computer technology and continuous systems supervision (black boxes), reporting has begun to be superseded by automated procedures. In this case, the initial difficulty involves "extracting the failure automatically from a stream of data". The real difficulty once this has been put in place will be knowing what to do with the no doubt impressive number of deviations catalogued by the automatic tracking system (out of all proportion to the number that are voluntarily reported by the actors today).

Automation of incident detection. A study carried out in 2010 [25] compares the results achieved using three methods used to catalogue serious adverse incidents (SAIs) in medicine in the United States: (1) a system of voluntary national reporting for medical professionals (the AHRQ or Agency for Healthcare Research and Quality system); (2) a system based on compulsory reporting by professionals of all incidents relating to a list of 20 national Patient Safety indicators (PSIs); and (3) an automated method for analysing the contents of all electronic medical records of patients admitted to hospital (global trigger tools method). The three methods were used for the same cohort of 795 patients from three general hospitals in 2004; the automated medical records monitoring system revealed 10 times more SAIs than the two other methods. A total of 393 SAIs were detected and 355 of these were only detected using the automated method.

The aviation industry has pioneered this process with its regulatory provisions for systematic analysis of on-board flight recorders (the tape for each flight is read and all abnormal values outside a normality envelope are subjected to additional manual analysis[10]). The analysis of medical records using the trigger tools method has been

[9] A good summary of this debate is found in Dekker [24].

[10] http://www.iata.org/ps/intelligence_statistics/pages/fda.aspx.

inspired by almost the same values [26, 27]: automated searches for abnormal values which trigger a subsequent manual analysis to understand the event.

The benefit offered by burdensome mapping methods is obvious for major industries but quite limited in innovative industries

The benefit offered by burdensome mapping methods (as compared with simple methods such as meetings between experts) is real, but ultimately it is quite fragile in terms of the time devoted to these formalities, particularly in highly innovative industrial systems.

To keep people's minds focused, a simple exchange of experiences over a few days involving a panel of carefully selected professionals who are carefully guided and actually work within the sector (guided brainstorming) will identify around 50 to 60 % of the total risk within a particular area; a feedback analysis (incident reporting) which is not carried out in depth and does not penetrate below the surface of anecdotes and immediate causes will yield almost nothing further (the stories that are reported are often tautological and confirm risks that are already known). On the other hand, an in-depth analysis of the same incidents yields 10 % more (i.e. 60–70 % of the real risk, once added to the initial brainstorming process) by identifying the systematic vulnerabilities (latent factors), but this takes a few days longer and experts once again have to be involved. Finally, formal methods (process analysis, functional analysis, FMECA, PRA etc.) add 15–25 % additional knowledge about the risk, but at the cost of a large investment of time (usually weeks or months). If all these steps have been carried out perfectly, which is rare and is only done in a small number of ultra-safe industries because of the time and resources required, the mapping results from combining the different methods may cover up to 90 to 95 % of the real risk. This is a very good result on paper but it has to be adjusted downwards to account for the natural obsolescence of the picture that results from it, which loses between 2 % (nuclear industry) and 20 % (medicine or software industries) of its relevance per year, depending on the pace of innovation and restructuring in the economic market for the system in question; due to the cost of the formal mapping process, it is very rarely conducted again at the same frequency…

Controlling safety in a context of high innovation: the case of medicine, with a knowledge turnover of 5.5 years.

Sjohania and his colleagues [28] analysed 100 reviews of questions published between 1995 and 2005 on recommended treatment strategies in multiple medical specialties, limiting themselves to the best randomised or semi-randomised controlled trials. They used two assessment criteria: quantitative, defined as whether or not a change in the clinical result by more than 50 % occurred in relation to at least one criterion as compared with the initial review, and qualitative, which considered efficacy, the identification of new complications or new gaps in knowledge that were not known at the time of the initial review. They found one of the two signals in 57 % of the reviews that were published. The average value for the half-life of knowledge before an obsolescence signal appears is 5.5 years. In 7 % of cases, the review already had an obsolescence signal at the time of its publication and 11 % of reviews had one within less than 2 years. Medicine is one of the few professional human activities with such a high rate of innovation (it is only really exceeded by the software industry). This frenetic pace of innovation is in contrast to the quality model that medicine has chosen to import, modelled on ultra-safe systems (particularly aviation), with a total time taken for deployment of the method for an innovative item extending to 10 years on average: two years to identify the problem, two years to come up with local solutions, one year to address it in a centralised way at the supervisory level, two years to come up with a consensual solution (a national recommendation) and two to three years to train all operators nationwide in the use of the solution. This is as good as saying that the quality cycle is never completed in medicine because of innovation. It will be realised that there is a need to take a special approach rather than using the methods that are known from industry, since innovation, even more than quality, is a force for progress in safety that no-one would wish to arrest—for example the shift to day surgery using minimally invasive surgical techniques and natural routes of access results in a fall in the number of hospital acquired infections (e.g. the abdominal wall is not breached, cross-infection is reduced to a minimum), while the methods proposed by quality approaches have had difficulty maintaining this (very high) rate for decades. This makes it clear that no health care professional will choose to reject this innovation and continue to rely on the quality effect, even if its introduction gives rise to other problems. What will be the use of maps created five years ago to address the risks of traditional surgery if we know that during the next five years a massive transition will take place to a different type of surgery in all Western countries?

The process of building defences after the results of mapping have been obtained is still a strategic node which is not easy to resolve

The last point (what safety strategy should be adopted, on the basis of what is known about risk mapping) is clearly even more strategic when it comes to action planning. That is the final point in this initial phase: specific risks are selected which it has been decided to protect, and barriers (defensive mechanisms) are then built against those risks. We should recall that there are three complementary types of barriers available [29] (prevention, recovery and attenuation barriers), and each of these makes use of a combination of intangible tools (training, laws and recommendations) and physical tools (failsafe systems, locking systems, access blocks etc.). We will now take a few moments to consider these points.

Selecting the risks being protected against: the relevant decision-making matrices

The mapping process provides a list of risks but no priorities for addressing them; a decision-making strategy is therefore needed which accepts certain risks and protects itself against others.

When it comes to the risk of accidents, the solution generally comes from using a frequency*consequence decision-making matrix.[11] The results of the mapping process are arranged in boxes along the frequency axis (from very frequent to exceptional) and along the consequence axis (from minor to disastrous). The decision-making process involves accepting certain risks (which occur very frequently but have no consequences, or which are extremely exceptional even though their consequences are disastrous) and protecting oneself against all the other risks. It is possible to protect oneself either by reviewing the permitted design or working conditions (prevention), or by increasing the capacity to recover from and attenuate risks (mostly through training).

This approach results in two risks: actually protecting oneself effectively from the risks that haec been identified as a priority and acknowledging the impasse situations that have been accepted. This last area is clearly the most difficult, and brings us back to the problem of dealing with the management of weak signals.

Weak signals are an attractive concept but one that often turns out to be illusory from a management perspective

The rationale described above leads to the building of defences against all those parts of the risk matrix that are judged to be acceptable to the system.

The weak point in this rationale is the risks that are excluded, particularly the weak signals that are often discussed in the literature and at conferences and which need to be listened to and taken into account more effectively [30, 32].[12]

This is because analysing weak signals means nothing other than analysing those parts of the current matrix that it has been decided not to analyse. What appears to be simple when expressed in this way actually turns out to be very complicated, for a number of reasons:

- the question of choosing what to include. Unlike the part of the matrix that one is dealing with, which corresponds to a finite number of elements, the part that is not dealt with or which is set aside, consists of an infinite number of elements (since it represents all the members of an infinite set minus those that are being addressed). It seems obvious that all our resources would not be enough to deal with an infinite set of potential risks; it is therefore necessary to make choices on what to include… but how? In this context, the problem of choosing what to

[11] In the area of production line quality, decision-making methods tend to be used that give priority to frequency (the Pareto method is the best known of these).

[12] Ostberg [31] RISCResearch Paper No. 3, http://www.wisdom.at/Publikation/pdf/RiskBerichte/RRR_GOestberg_SomeIntangibles_09.pdf.

include comes back to the accident model that is developed to deal with these events. This choice in turn can be analysed into two other questions;

- the question of the accident model that is chosen. Weak signals cannot be dealt with using standard accident models, because it is precisely these (weak) signals which are (quite reasonably) set aside because they are not sufficiently severe or frequent. It is necessary to use models using percolation or coinciding conditions that find minor signals and events when they coexist in the same context; when these signals occur together, this constitutes a risk event. It goes without saying that the management and complexity of these percolation models has nothing in common with the simple models; these aspects take time, consume resources (the processing of weak signals that are rejected creates a considerable additional workload) and above all a high level of competency in the underlying field (the qualified individuals are usually university staff and are rarely employed in enterprises);
- the third limiting factor is the macroeconomic cost of extending monitoring to include weak signals. Thanks to a number of studies carried out in various places, the additional cost can be estimated to be 5 to 10 times the current cost of safety [33]. There are two types of added cost: of course the signals have to be included in the analysis, but there are also the indirect costs resulting from the protection strategies that would be developed to combat these low risks (which would cause some industrial initiatives and innovative risk-taking to grind to a halt) [34].

Although the concept of weak signals is very attractive in theory, it is easy to establish that it is quite unrealistic when it comes to the Cartesian management of everyday risks.

Fundamentally, weak signals have their greatest social utility through the actions of whistle-blowers. This existence of a counterbalancing power and a form of activism that asks questions about risks that are rejected or neglected, even if no in-depth analysis occurs of the Cartesian fundamentals in response to their revelations, at least maintains the feeling in society and among risk managers that not everything—indeed far from everything—is being controlled by the risk and accident models that are worked out on rational grounds.

Step 2: Comparing the Paper Model with the Reality on the Ground

Once the theoretical defence model has been built on the basis of the mapping process, it has to be put into practice and made sustainable: theories are called into question by real experiences and this raises discrepancies which have to be understood and either refuted (or accepted and changes made to the model) for the model to retain its relevance.

The most difficult problem to deal with is that of the progressive migration of the system and the automatic increase in violations as safety is improved [35].

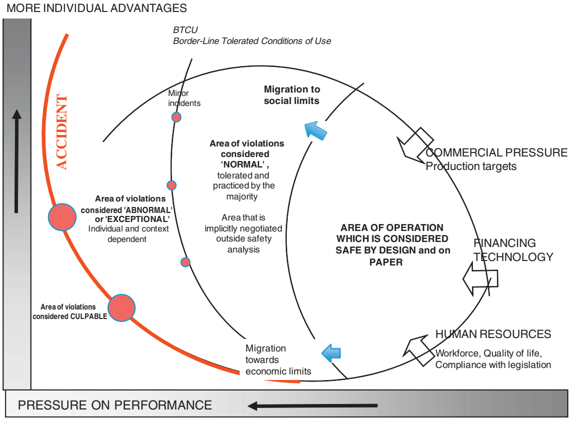
Migration of practices is the norm in all systems [36]. Technical and economic conditions regularly impose new constraints on working situations: in many cases more (performance) has to be done with less (personnel, resources). Such degraded conditions, which initially only arise occasionally or during critical commercial periods, are not immediately punished by poor results or incidents (generally the opposite happens, as positive results are achieved thanks to the increase in performance); as a result the migration becomes standard and is accepted by everyone. This "illegal-normal" standard is accompanied by beneficial feedback (for workers) and this is perceived as corresponding to and compensating them for their efforts in terms of production: the hierarchy often gives them greater autonomy (specific initiatives are contested less, timetables and replacements are allowed to organise themselves, in short more is tolerated, bonuses are awarded for severe staff shortages and these deviations are gradually omitted from the feedback process). It would create problems if these migrations did generate signals, because they are providing a service both to the enterprise (performance) and to the workers (secondary benefits). The proportion of the enterprise that is operating within an "illegal-normal" range can easily rise to as high as 40–50 % of existing procedures in systems that are under economic pressure.

The size of this hidden area that does not conform to the prescribed model, is proportional to the appearance on paper of the margin that was built into create safety buffers: the violation or deviation are in fact only defined in relation to a regulatory requirement (whether this is inside or outside the enterprise itself). If the regulatory requirement contradicts an increasing demand for performance, the number of violations will automatically increase and the system will begin to migrate.

Two paradoxical points to remember about violations:

(a) violations are a characteristic of safe systems (which have procedures); they do not exist in activities that have no legislative framework and no rules;
(b) procedures that are poorly designed and too demanding automatically create violations.

A typical model of migration in practice (Amalberti et al. [36, 37], inspired by Rasmussen [38]): professional practices are controlled and limited from their conception by a combination of rules and formal and informal barriers to avoid the pressure of production causing rapid migration towards a high-risk area. These pressures on production, however, are so strong that the system will migrate, particularly if the system has been locked into very (excessively) prudent practices since it was first designed. Practices will migrate towards greater productivity and also towards greater benefits for operators. These "normal" violations, which are tolerated by everyone, may affect up to 50 % of procedures and in some cases will continue to be amplified, creating real risks. It is the risk of secondary migrations that constitutes the priority target for interventions in safety.


MORE INDIVIDUAL ADVANTAGES
BTCU
Border-Line Tolerated Conditions of Use
ACCIDENT
Minor incidents
Migration to social limits
Area of violations considered 'NORMAL', tolerated and practiced by the majority
Area that is implicitly negotiated outside safety analysis
Area of violations considered 'ABNORMAL' or 'EXCEPTIONAL' Individual and context dependent
AREA OF OPERATION WHICH IS CONSIDERED SAFE BY DESIGN and on PAPER
COMMERCIAL PRESSURE
Production targets
FINANCING
TECHNOLOGY
Area of violations considered CULPABLE
Migration towards economic limits
HUMAN RESOURCES
Workforce, Quality of life, Compliance with legislation
PRESSURE ON PERFORMANCE

How can migration be controlled? There are three solutions available to specialists:

- the first is a good idea which is nevertheless false, in relation to training: it involves front-running migration through unlimited increases in the skills of professionals so that they are able to deal with the exceptional peaks in production demand but return to conformity mode once the peaks have passed. As we saw in the last chapter, this solution contradicts the real situation: the more operators are technically trained and become even more expert, the more they will integrate their past successes, become confident and continue to migrate "one step further" even under normal conditions;
- the second is typically systemic and relates to design. We have just seen that more violations take place the more ideal and unrealistic the safety model imagined in the design on paper has become and the more it fails to take into account the economic constraints on production. At this stage we should remember that it is better to have a safety model which is designed to be less ambitious and compatible with the required performance rather than the other way round. We must also remember two other points: no safety model can absorb all the future changes that will occur in economic conditions from the time when it is designed. It is therefore quite normal for a model of practices to migrate in order to adapt to these new economic conditions, and it is therefore also normal to adapt the safety model regularly, sometimes relaxing its constraints and not necessarily systematically strengthening them. Worse still, since all violations and all migrations are a sign of dissociation between the requirements of the safety model and the performance model, the response to migration beginning must never be to reinforce the procedure or the constraint, since such a response will enlarge the difference between what is prescribed and what actually happens and will automatically increase the number of violations;

- the third is typically sociodynamic, and concerns monitoring of individuals who deviate more than others: migration of practices towards the illegal-normal opens up possible breaches in terms of social tolerance of non-compliance with the procedure [39]. If nothing is done, certain individuals who are more prone than others to take an autonomous stance will quickly extend these local permissions to all their practices. These individuals can never be controlled by orders from the hierarchy, at least not if they have had no accidents; worse still, they are often brilliant individuals whose accumulated successes and performance records are often presented as examples and who hold an envied status within the group. The only way to contain these individuals is to put them under the "authority of the group", an old historic solution inspired by the functional protection that groups and families provide to the most fragile individuals, and which is often found today in modern ethno-psychiatry. Setting up a reciprocal, consensual surveillance system between the members of the group, with the involvement of local managers and under their supervision, has proven to be highly effective in reducing extreme forms of individual deviation (but of course not those forms that have become a consensus for the whole group). It is specifically these techniques that underlie the success of "sellers of regulatory conformity for safety at work", who promote the techniques of BBS (Behaviour-Based Safety) such as the DuPont method. These methods are used in non-typical ways to reduce deviations from rules on wearing safety equipment in industry [40]. Nevertheless, making these techniques work requires a powerful team dynamic (all deviations have to be mutually reported, this has to be done politely, time must be taken to understand why etc.); this means more time and more personal involvement, as well as a renunciation of the cult of success and hero-worship. It is not surprising that these techniques are rarely introduced, except in the limited case of wearing safety equipment in factories, and if they are introduced they also turn out to be very difficult to maintain over time.

Step 3: Widening the Angle of Attack, Addressing Safety Through Macro-Scale Organisation

Making the operator's immediate environment safe is not enough to make a whole system safe. The first two steps are almost exclusively located at the micro level (the workplace and the area nearby) and at the meso level (the nearby enterprise).

A system consists of much more than this.

The general economic conditions governing trades, professions and regulators are very important when it comes to determining the existing opportunities to make the system safe [41].

A difference in safety by a factor of 100 to 1000 between the safest and least safe systems

The differences between safety levels in industry rise to a factor of 100 in terms of safety at work and a factor of 1,000 in terms of process and plant safety. For example, in the area of safety at work, 1 in 1,000 professional fishing skippers die at work every year, as compared with 1.5 deaths per year for every 10,000 workers in the construction industry, and less than 5 deaths per year for every 100,000 workers in all other sectors [42]. The safety figures for industrial installations and processes are even more encouraging. One in 1,000 patients dies from an unexpected complication associated with an unwanted effect of his treatment in hospital, as compared to a risk of 1 premature death (associated with risk exposure) per million residents living for 5 years in the immediate neighbourhood of a nuclear power station.[13] or one premature death per million passengers who take an airline flight.

The effect of these differences is that there is little capacity to make relative changes in the level of safety between the major classes of human activity. For a change to be visible on this scale, it would be necessary for medicine, for example, to improve safety by a factor of 10 (which is a real challenge, since it has achieved virtually no progress in 10 years), and even such a change would still be negligible in comparison with systems such as civil aviation, which are still 1,000 times safer.

The overriding importance of macro-scale challenges in relation to local actions to improve safety

The systemic factors and central management of the enterprise often impose major limitations in terms of safety initiatives that can be pursued at the local level [38].

As a minimum, these may involve a willingness to coordinate all policies centrally while putting a brake on isolated local initiatives; more frequently still, the management are aiming to preserve the contradictory tensions between the real need for safety, to maintain their licenses by responding in accordance with administrative requirements imposed by regulators, while protecting the desire for exposure to acceptable risks in order to strengthen the image, maximise production, support the economic model or promote the personal success of managers.

Safety specialists have to learn to deal with this paradoxical framework and understand the trade-off mechanisms that are used in their industry.

The nuclear industry, for example, has a very low social tolerance of voluntary exposure to risk. The same is true in civil aviation, so here the trade-offs tend to be made in favour of safety initiatives, while targeting full and centralised coherence (and working to prevent isolated local actions). The same has not been true (at least until recently) in international finance, medicine, fishing and motoring, to cite just a few examples; in these cases, senior management—or the institution in the broad sense—gives priority to exposure to risk and safety activities are mostly conducted at the local level, and it is accepted that they are local and limited in scope.

It is understood that a safety specialist may move from one system to another during his career, using the same tools and the same knowledge and yet obtaining very different results. The context in terms of risk acceptance serves to buttress the level of safety, while local actions are only flimsy supports, whose leverage has a

[13] Wilson [43], or http://mullerlbl.gov/teaching/physics10/old%20physics%2010/physics%2010%20notes/Risk.html.

very limited duration and geographical scope and which are less likely to be rolled out by the management the more they are perceived as incompatible or detrimental to the overall economics of the system.

We will come back to these points later on when we analyse the work that has been done on types of safety and margins of action, and also on ways of building trade-offs and compromises.

Five barriers that account for the differences between less safe and ultra-safe systems [44].

The first barrier is the lack of any limitation on performance imposed by the authorities regulating the system. Fishing skippers at sea, for example, are able to remain at sea even in force 10 conditions, professional mountaineers may decide to embark on an attempt on the summit regardless of the risk conditions, and the on-call doctor in the emergency team has to see every patient who comes in, with no hope of relief even if he is exhausted. In the absence of regulations that act as a brake on exposure to risk, these professionals cannot hope to have a safety level higher than 10-3. The second barrier involves the autonomy of the actors themselves. If the only factor that matters to you is the realisation of your personal goal, and you do not take into account the consequences for those who are following after you, there is no chance of that you will reach a safety level that is higher than 10-4 for your activity. A surgeon who is running 4 h late with his list will be able to complete his list without making a mistake, but the risk will result from a patient who has had an operation returning to the ward at midnight or 1 am, in conditions where there are fewer nursing staff and at a time that creates greater risks in terms of post-operative monitoring. A safe approach would require him to take into account the limitations on the ward and no doubt to cancel his surgical list once it had become delayed in this way. The third barrier "or the artisan's barrier" summarises the effects above in social terms. One does not ask the name of the airline pilot or the biologist, but one chooses a jeweller or a surgeon; they are artisans; they market their differences and their personal knowledge. There is no skilled trade in the world, however, with a level of safety higher than 10-4. The reason for this is simple: safety is based on reducing differences and maintaining a stable service 24 h a day, while the artisan markets his services precisely on the basis of the difference between himself and his neighbour and competitor: the instability of the system as a whole is therefore structurally inherent in his work. The fourth barrier, which is referred to as "overprotection", occurs in systems that are already safe. The safer the system becomes, the more paradoxically exposed it is to enquiries and to blaming of those responsible whenever an accident does happen; the response is often to create rapid growth in protective procedures that are intended to provide cover for management. These protective procedures place unnecessary restrictions on work, make it more burdensome and promote deviation and migration of working practices among front line operators, while paradoxically increasing the risks that they are intended to protect against. Finally, the fifth barrier relates to the loss of rationality in communication about safety in systems that have become very safe. Fewer accidents occur, but they have a higher media profile. One important component of communication by managers and the actions that are taken therefore involves reassuring the media. Statements and assurances are often issued precipitately (there is a tendency to read the surface-level statistics and boast about progress without taking a step back) along the lines that the press wishes to hear (and not necessarily along the lines of the safety model). Together with this approach of making decisions at multiple levels with no assessment of their objective results, the safety system becomes a huge, multi-layered system of procedures and requirements, where nobody knows which ones actually contribute towards achieving the level of safety; as a result the system becomes incapable of self-cleaning, becomes excessively complex and creates an asymptotic trend in safety [45].

Step 4: Once All the Steps Have Been Taken, is it Still Possible to Withstand Exceptional Events?

Managed safety and controlled safety

The safety of complex systems is the sum of two aspects: on the one hand the safety achieved through all the prohibitions, limitations, legal requirements (referred to as controlled safety) and on the other, the safety supported by the adaptive intelligence of the operators and professionals within the system (managed safety). These two concepts, which are now widely discussed in the literature, were introduced for the first time in 2008 in the leading article published in Human Factors on the work done on professional fishing skippers in the thesis by Morel et al. [6].

Total safety = controlled safety + managed safety

Let us consider the terms of this equation so that we can apply it to different fields of professional work.

Skilled trades are subject to few regulations; their whole of their rather modest safety structure is based mostly on the quality and skills of the operators themselves, with the high level of variation that is inherent in individual quality. On the other hand, for these experts adaptation to exceptional conditions is their everyday work and they are remarkably good at it (at least this is true of the best of them, who survive both economically and physically).

Total safety (skilled trade systems) = controlled safety + managed safety
(*the font size metaphorically shows the strength of each term in the equation*)

Very safe systems, on the other hand, have large numbers of procedures and prohibitions. The level of safety is high, but the adaptive expertise of operators in these areas is automatically reduced because they are no longer exposed to exceptional situations and they are no longer trained to work outside their procedural framework.

Total safety (ultra-safe systems) = controlled safety + managed safety

There is currently no known solution that can preserve both the expertise of the operators in exceptional situations and the benefit of achieving maximum system safety by procedural means.

It should be remembered that resilience, or the art of adapting to exceptional conditions, is a native feature of human systems that rely on their own autonomy to survive; it automatically disappears as a result of using the traditional tools that enhance safety in industry and service sectors. In many cases it would be preferable to have less of this in less safe systems (because it is associated with frequent improvisation and habitual non-compliance) and also to reintroduce it in ultra-safe systems (because it would permit adaptation to exceptional situations, an ability which by this stage has largely been lost).

The safety choice in the aviation industry: suppressing heroes and prohibiting training in situations that are too exceptional.
In 1995 two serious accidents and incidents occurred involving Airbus A310 aircraft which suggested that pilots do not have enough training in difficult manoeuvres. The first was an accident which occurred on takeoff from Bucharest, where the crew were distracted or busy and the aircraft was allowed to go into overbank and was unable to recover from this unfamiliar situation; a few months later, when coming into land in Paris, the crew of another Airbus A310 were surprised to find the aircraft in an unexpected attitude and it was only with great difficulty and considerable good luck that they recovered the aircraft without damage. The two inquiry commissions emphasised the lack of training given to these pilots who fly modern aircraft but who are clearly better trained to manage the computer and use the automatic pilot than to fly manually themselves. The response from the civil aviation authorities was uncompromising: there was no question of resuming training in manual flying, which only occurs in exceptional circumstances. This would re-open the door to the "heroes" who depart from standard procedures too frequently; the aviation industry has worked long and hard to eliminate them. The chosen solution will be to aim to use warning signals to provide better indications that the aircraft is moving outside its normal flight path, and to rely more on automatic systems and on the safety net to recover from these unusual situations automatically. This attitude has been seen in relation to every single unusual accident.

This is an exaggerated example of a system (civil aviation) which relies completely on procedures and supervision and which benefits from these every day in terms of safety results (one of the best safety standards in the world), but is now locked into that approach. Any attempt to reintroduce improvisation in areas where there are no procedures is initially seen as putting the system at risk and is consequently not allowed.

It should also be noted that the very well-known case of the successful landing on the Hudson River by the US Airways flight in 2009 is simply viewed as a stroke of luck in the world of aviation regulation and is not leading to any questioning of the model that has been described above (there has been no collective and professional learning experience from this accident in the aviation sector, even though it has supported a number of research groups who wish to re-use it for their own purposes).

Institutional resilience: surviving an accident can prove to be as important as avoiding the accident

Safety specialists have a simple and unique model in their minds: reduce the number of accidents and incidents. This means forgetting a whole aspect of safety which is typically systematic: knowing how to survive accidents.

This approach may appear to be cynical, but it should be integrated as a fully-fledged component of the modern approach to safety in an enterprise or complex system.

If this is done, it adds a critical aspect to the two well-known themes of crisis management (knowing how to respond to the accident and knowing how to manage communication) by emphasising "the ability to continue production after the accident". When considered more closely, this third aspect may even be the most important of all. It encourages the continuation of the crisis cell long after the accident, to manage "long-term damage to the image".

The accident is not just an isolated incident in the life of a system. It may be (and often is) repeated at regular intervals. Every accident is like a heart attack or a relapsing cancer for a human being. It calls into the question the survival of the whole system.

A model of institutional resilience: the example of Air France.

The airline Air France has suffered five major accidents in 15 years (747 in Papeete in 1993, 747 cargo plane in Madras in 1999, Concorde in Paris in 2000, A340 in Toronto in 2005, and A330 between Rio and Paris in 2009) and a series of major incidents during the same period. This makes Air France the riskiest major airline in the world. What is more, unlike its competitors which have had fewer accidents (in some cases only one) leading (in almost all cases) to their economic demise (examples include TWA and Swiss Air), Air France managed to save its image from being destroyed after these accidents and in some cases even gained market share! This is an example of resilience on the part of the enterprise. One example: on the night in 2005 when its A340 crashed at Toronto airport as a result of poor judgement by its crew, the Air France management successfully created a positive reading of its company image, both immediately and in the longer term, thanks to a "positive" angle on the accident which was broadcast by all press outlets and international news agencies, since it demonstrated the airline's ability to evacuate an aircraft in difficulty. This was the abiding image in the public mind. The ability to do this owes nothing to chance. It is prepared in advance, is managed directly and forms part of a highly rehearsed post-accident plan. This ability is part of the company's know-how.

Three Models of Balanced Safety Rather Than Just One

The idea of a single model of safety that applies in every context and aims to have zero accident is naïve. Safety is a social construct and it adapts to demand. As this section shows, there are many different responses to safety, which describe a number of different models of safety (resilience, HRO, ultrasafe), each with their own approach, advantages and limitations. These models take different approaches to the trade-off between the benefits of adaptability and the benefits of the level of safety.

Three Very Different Strategies in Terms of Exposure to Risk

Everyone will agree that writing a safety plan offers no guarantee that the plan will be put into practice.

Everyone will also agree that it is rare for a safety plan to mention what is not going to be done because of the trade-offs that have been agreed.

These two aspects (doing well what one has decided to do, and knowing what one has decided not to do) are strategic in terms of managing risk on the ground.

The compulsory pencil and paper risk mapping exercises and the risk prevention protocols developed for regulators are no doubt essential, but these are often only one-time efforts (for demonstration purposes). They need to be rooted in everyday reality over a long period of time… which presents a completely different problem.

In the absence of concerted action and trade-offs at a higher level, front-line management and front-line staff will be the level at which all the (contradictory) constraints issued by the various divisions are integrated: produce more, under conditions requiring quite high levels of situational adaptation and in accordance with the safety instructions and rules. There is every reason to fear erroneous local interpretations and new vulnerabilities.

Three families of safety plans or solutions are always available to sociotechnical systems for the management of everyday risks.

- **PLAN A. The first family of safety solutions involves eliminating or delaying exposures to risk. We will call this plan A**: the aviation industry excels at this strategy. Thanks to its global coverage and its absolute authority, air traffic control is able to prevent situations where aircraft are exposed to difficult conditions. The same applies to the nuclear industry, which has very robust incident procedures, all of which tend to ensure the immediate safety of the installation and shut it down temporarily. In return, this level of supervision promotes economies of scale in training: there is no point training pilots to fly aircraft in hurricane conditions if one knows it is possible to avoid all hurricanes. This prudent strategy, however, also requires a high standard of systematic supervision which is often outside the scope of fragmented, deregulated and/or highly competitive industrial activities and therefore also of skilled trade systems.

The ability to implement plan A depends on the way the system is organised. Let us take the example of a comparison between two health care systems in France and the United Kingdom, where a hip replacement is carried out for a patient with comorbidities (diseases other than the hip problem, such as diabetes or hypertension). In these cases the replacement operation is never an emergency procedure and it is better to wait for conditions in which the comorbidities are perfectly under control before the operation takes place, to avoid postoperative complications. This strategy of waiting for favourable conditions works quite well in the United Kingdom, because the system of access to health care is highly controlled; on the other hand it is quite ineffective in France, because there is too much private provision and too little regulation; patients can consult as many surgeons as they wish and these consultations are reimbursed, allowing them to get an operation date very close to what is convenient for them. It is not surprising that in this context French surgeons take more risks than their English equivalents to avoid the patient leaving, and use plan A less often.

- **PLAN B. The second family of safety solutions involves accepting exposure to risk while complying with all the recommended standards and procedures (we will call this plan B)** but while maintaining the ability to detect changes in the context and maintaining local, intelligent adaptation of these procedures (the procedure at the heart of the system). Strict and standardised implementation of all professional recommendations under standard working conditions will minimise the number of accidents. These plan B and plan A approaches generally feed into the responses that are made to regulators.

There is no such thing as a working environment without incidents and above all without surprises. In order to gain the maximum benefit from this approach, it is therefore necessary either to be able to stop the system and make it safe quickly (no go), or to have unambiguous procedures for dealing with incidents (specific procedures to deal with each catalogued type of abnormal event). In this case the operator does have to be aware of how the situation is evolving and must be able to identify the problems and apply these procedures.

It should be noted that poor system ergonomics can easily compromise this aim. Dave Woods, setting out the pioneering work by Bainbridge [46], has extensively defined the risk of “surprises” in standardised situations in ultra-safe environments where changes do not occur often; such surprises are mostly associated with designs and supervisory systems that do not incorporate the ethical rules to ensure the stability of the tool being used (see the example below). From this, Woods and Hollnagel [47] deducted a number of principles of ergonomics that were to characterise the design of safe systems; in particular they reiterated, and popularised under the term Joint Cognitive System the old idea that had been put forward by French ergonomists since the 1960s[14]: “ergonomic analysis is on the wrong track when it breaks down the work in a Cartesian way and analyses it in separate parts (work analysis)”; the activity of the operator is embodied in the technical context, and can only be studied in the form of a dynamic linkage. The contribution by Mica Endsley should also be noted [48, 49], which follows the same track and proposes testing the adaptation of the understanding of the world within which the operator develops, using his concept of *Situation Awareness*.

One example of an unacceptable surprise associated with a design [50]. In the late 1980s, the first Boeing 737 s had an ALT HOLD function which made it possible to stabilise the altitude of the aircraft in the vertical plane (stop it from climbing or descending) by simply pressing a specific button; pressing this button had no effect during the final phase before landing, however, because the aircraft builder wanted to safeguard against accidental pressing of the ALT HOLD button, which could have seriously disturbed the automatic landing process under conditions of poor visibility (CATegory 3 procedure). It should be noted that this protection system had been installed to meet the safety requirements imposed by the regulators to obtain the CAT 3 qualification. The result was telling: during this final phase, to achieve the same result (stabilise altitude) a complex sequence of actions had to be carried out: first disconnect the automatic pilot, then disconnect the two flight controllers at the controls on the right and on the left, then reconnect the automatic pilot, and finally press ALT HOLD. A number of accidents and severe incidents were caused by this ergonomic inconvenience before it was corrected: pilots, not realising that they were in these approach conditions in which direct use of ALT HOLD

[14] deMontmollin M. L’ergonomie de la tâche, Peter Lang, Berne, 1986.

was prohibited, pressed the ALT HOLD button and nothing happened, and this was followed by a moment of surprise and inappropriate action, which caused incidents/accidents.

- **Plan C. The third family of safety solutions involves tolerating exposure to non-standard conditions, while accepting that operators do improvise or deviate from procedural behaviour. We call this plan C.** In many professions, life does not consist of procedural repetition; quite the opposite. This capacity for adaptation has led to a great deal of debate in the small community within the humanities in which the idea of resilience is studied.

The Metaphor of the climber and the rock face. One can consider **hazards** as rock faces. They are an inevitable part of nature. In industry, such rock faces may represent sick patients, the chemical properties of compounds, solar radiation etc. Risks depend on the willingness to deal with these rock faces and the way in which this is done. One can refuse to climb them (plan A), one can limit oneself to climbing only known rock faces and follow all the required procedures (plan B), or one can attempt rock faces in non-standard situations (without equipment, without training, under poor or changing conditions), or worse still, climb unknown rock faces (plan C). The more stable and supervised it is, the more it relies on avoidance, plans A and B, and the less stable it is, the more it will have to rely on its adaptability to deal with changing conditions (plan C).

In short, these solutions cannot be transferred from one environment to another; they have different aims, use distinct models of safety, require different types of experience and training and follow different organisational approaches.

Outside a small number of ultra-safe industries, the majority of human professional activities rely heavily on plan C. Strangely enough, however, all the literature on the quality and safety of systems offers prescriptions only for plans A and B.

It is not because those relying on plan C do not follow all the procedures and result in improvisation that it is not possible to make their practices safe. The problem is that the solutions that would make these practices safe while accepting their reality do not consist in developing procedures. (If they did, one would change to a plan B approach. Instead, the response is ad-hoc and does not cover all the situations that arise during the work, whose very economic rationale often demands that it rely on plan C.) Plan C solutions are found in quite resilient models: becoming more expert, becoming able to judge the difficulty of the task according to one's own skills, learning to learn, drawing from experience, acquiring generic knowledge schemas which allow adaptation to borderline circumstances.

Three generic plans to manage risk:

- PLAN A: refuse to do it or wait for ideal conditions;
- PLAN B: do the work under ideal conditions according to recommended procedures;
- PLAN C: accept that you have to take action without having ideal conditions, including improvisation and working outside procedures.

In aviation, the ratio is 40 % plan A, 55 % plan B and 2 % to 5 % plan C.

In medicine, the ratio is 5 % to 10 % plan A, 40 % plan B and 55 % plan C.

What is the ratio in your own work between plan A, plan B and plan C?

If your plan C ratio is higher than 5 %, what is the value of the procedures that you have in place to cover plan B in the cases when you are working according to plan C?

Two professional contexts and two diametrically opposed safety strategies = helping to survive the risk versus protecting operators against exposure to risk.

Systems that have a relatively modest level of safety (lower than 10^{-4}) have considerable exposure to risk because they literally make a living from that exposure. This is true of fighter pilots, sea fishing skippers and professional mountaineers. In these professions, accepting exposure to risk and even seeking out risk forms the essence of their work. These professions do, however, still want to improve safety. A number of studies carried out among fighter pilots [51] and sea fishing skippers [8, 52] show a real need for safety. Fishing skippers, for example, would like to have an intelligent anti-collision system to offer them better protection in high seas with poor visibility and with the mobility required for trawling (Automatic Radar Plotting Aid). Fighter pilots would like an electronic safety net to offer them better protection when they are undertaking manoeuvres that are likely to make them lose consciousness (Electronic Safety Net). Moving on to the example of civil aviation, everyone also wishes to improve safety in this area. Here, however, the solution is radically different and most commonly involves not exposing crews to the surprising conditions or risks that are thought to be the cause of accidents. For example, the eruption of the Eyjafjallajökull volcano in Iceland in 2010 led to all aircraft immediately being grounded, based on a simple approach: no exposure to risk. These different examples highlight two completely opposite strategies to dealing with risk: one, which is supported by small-scale systems involving skilled trades or highly competitive activities, involves relying on the intelligence of operators and giving them aids to deal with risk; the other involves relying on the organisation and supervision and ensuring that operators are not exposed to risks. It is easy to understand that both of these models have their own approach, but in that case it is also necessary to accept that the safety solutions are not identical in both cases.

Three Authentic Models of Safety Rather Than Only One

Taking into account the risk exposure strategies already mentioned, it make sense to take the view that each one has given rise to an authentic way of organising safety which is original, with its own approach and its own possibilities for improvement [53].

- **The resilient model** involves professions in which seeking exposure to risk is inherent in the economic model of that profession. Skilled trade professions in particular sell their services on the basis of their expertise which allows them to deal with new risks, or even deal with the unknown, innovating, mastering new contexts, coping, winning through and reaping benefits where others fail or are afraid to go. This is the culture of champions, winners… and losers (the losers are part of the context, but they are not perceived as failures of the system but rather as a reflection of the knowledge and skill of the champions). Sea fishing skippers, for example, are capable of seeking out the riskiest conditions in order to prioritise catching the most profitable fish at the best times (sales economy); experts in oil exploration have to find oil almost at all costs once the procedure has been initiated; only success makes sense at that point. Traders constantly have to maximise their profits and military fighter pilots[15] always have to win… All these professions have objective accident statistics which are more or less disastrous. They are not, however, insensitive to their professional risks, and they deal with these through safety and training strategies which are very well thought-out, but of course within a different culture.

In these professions, the individuals' autonomy and expertise take precedence over the hierarchical organisation of the group. In many cases the group is very small (consisting of two to eight individuals) and works in a highly competitive setting. The boss is recognised for his technical ability, his past performance and

[15] The case of fighter pilots is a special and interesting case of a dual context: in peacetime, their administration (the Air Force) operates essentially on an ultra-safe model, but once the aircraft are deployed on active service, the operating model suddenly changes and returns to its fundamentals of resilience. These very contrasting contexts do generate surprises in terms of safety in both directions: persistence of resilient, deviant behaviour (as compared with the model that would be desired in peacetime) after returning from military campaigns, and important opportunities that are missed during the first few days of engagement due to lack of practice in the resilient model, when pilots are suddenly thrust from peacetime into operational theatres. A military air force can also shift the crew of an AWACS surveillance aircraft from peacetime into wartime during the space of a single mission: they may leave a French base in France, having dropped their children off at school in the morning… fly a 12 h mission that involves working in and overflying an operational theatre with a very high risk of aerial engagement and requiring particularly high resilience and return the following night to their air force base in France and also to their homes, which are completely organised around social routines and the challenges of peacetime.

his charisma more than for his official status. Every operator is constantly invited to use a very wide margin of initiative. A correct assessment of his own skill, courage and accumulated experience are the keys to recognition as “a good professional and a winner”; safety is mostly about winning, surviving, and only winners have a chance to communicate their safety expertise in the form of champions’ stories. To summarise, there are a small number of procedures, a very high level of autonomy and a very large number of accidents. It is still possible to make progress in terms of local safety, however, by becoming better trained through contact with the best masters, learning from their experiences and adding to one’s own mental capacity to adapt to even the most difficult situations. The differences between the least safe and the safest operators within a single resilient, skilled trade are of the order of a factor of ten,[16] which proves that it is possible to make progress through safety interventions, even while remaining within the “micro-Gaussian” distribution of professionals engaged in these hazardous types of work.

- **The HRO model** (*High Reliability Organizations*) uses the same idea of resilience, since it also promotes adaptation, but this is a kind of adaptation which is more local and controlled, involving human activities which are clearly better organised, with less of a tendency to seek out daring exploits (which is more characteristic of the pure resilient model). The HRO model is in fact relatively averse to individual exploits that are not controlled by the group.

HROs typically apply to professions in which risk management is a daily affair, even if the aim is still to keep risk under control and avoid unnecessary exposure to it: firefighters, merchant navy and naval armed forces, professionals in the operating theatre, oil exploration, those operating chemical factories.

HROs rely on the leader and the professional group, which incorporates several different roles and types of expertise in order to maintain a constant perspective on progress being made towards the goal (while avoiding the risks of a local focus), where all the members of the group play a part in detecting abnormalities in a contextual setting (sense making), bringing them to the attention of the group, adapting the procedure to these changes in the context. This includes deviations from procedures when necessary (but only when this makes sense within the group and is communicated to everyone). All members of the group show solidarity in terms of this safety objective.

Combating adversity is an integral part of the HRO approach but the high level of collective regulation (not necessarily only by the leader) imposes considerable limitations on isolated individual initiatives and promotes prudent collective decision-making.

The HRO model analyses its own failures and seeks to understand the reasons behind them. The lessons drawn from these accident analyses, however, are

[16] The rate of fatal accidents in professional deep-sea fishing varies by a factor of 4 between ship owners in France and by a factor of 9 at the global level, source: Morel et al. [52], op. cit.

primarily about ways in which the situation has been managed and could be managed better in future.

This is therefore a model which relies firstly on improving the barriers to detection and recovery, and secondly on barriers to prevention (which involve not accepting exposure to these difficult situations). Training is based on collective acquisition of experience. Once again, the differences between the best operators and those that are less good within a single trade are of the order of a factor of ten.[17]

- **The ultra-safe systems model** no longer makes it a priority to rely on the exceptional expertise of these front-line operators to escape from difficult situations; instead it requires operators to be identical and interchangeable within their respective roles, and in this case requires them to work at a standard level. The model, on the other hand, relies upon the quality of external supervision, making it possible to avoid situations where these operators are exposed to the most exceptional risks; by limiting the exposure of operators to a finite list of breakdowns and difficult situations, the model can become completely procedural, both when working under normal conditions and under abnormal conditions. Airlines, the nuclear industry, medical biology and radiotherapy are all excellent examples of this category. Accidents are analysed to find and eliminate the causes so that exposure to these risky conditions can be reduced or eliminated in the future. This model relies on prevention first. Training of front-line operators is focused on respect for their various roles, the way they work together to implement procedures and how they resopnd to abnormal situations in order to initiate ad-hoc procedures. Once again, the best and the least good operators within a single profession differ by about a factor of ten.[18]

Four lessons can be drawn from this:

- **the three models of safety are radically different**. They represent responses to different economic conditions, each one has its own approach to optimisation, its own approach to training, its own advantages and its own limitations. They can be plotted along a curve in which there is a trade-of between flexibility and adaptability on the one hand, and safety on the other. All three, however, have the same capacity for internal self-improvement, and safety can be improved by a factor of 10 (making them 10 times safer);
- **the three models cannot be mixed**. Mixing the features of one with those of another leads to a failure to improve safety and may even be counterproductive.

[17] The rate of fatal industrial accidents in the gas and oil extraction industry varies from 130 deaths per 100,000 workers in some African countries to 12 deaths per 100,000 workers for the best oil wells; the global average is 30.5 deaths per 100,000 workers, source: http://nextbigfuture.com/2011/03/oil-and-gas-extraction-accidents-and.html.

[18] The rate of aviation accidents ranges from 0.63 per million departures in Western countries to 7.41 per million departures in African countries. These therefore differ by a factor of 12, source: IATA statistics, 23 February 2011, http://www.iata.org/pressroom/pr/pages/2011-02-23-01.aspx.

For example, there is no certainty that reintroducing training in deviating from procedures and dealing with unknown situations in civil aviation will not erode safety rather than improving it (this is why global regulatory authorities have refused to go down this route). On the other hand, introducing restrictive procedures in combat aviation or deep-sea fishing would perhaps result in fewer deaths… but it would kill off the profession itself;

- **local interventions cannot change the model**. If intervention takes place locally to improve the safety of an enterprise or a particular working unit within a profession that belongs to a specific model (resilient, HRO or ultra-safe), there is no opportunity to encourage the adoption of characteristics of a different safety model (for example, if intervention takes place in fisheries, it would represent an illusion to advise them to adopt an ultra-safe system strategy). Instead, it is necessary to rely on the capacity for progress that is available within the model in that specific professional setting (the resilient model, in the case of fisheries) by using the strategies that are specific to that field and have been seen to offer significant opportunities to improve safety, since this can be improved by a factor of 10;
- **it is possible to switch from one model to another, but this requires a changeover event** that will affect the entire profession and its economy. The industrial chemical industry, for example, which in some cases is still based on resilient models dating from the 1960s and 1970s, made a definitive switch to an HRO model after the events that occurred in Seveso in Italy in 1976 and the European Directive that followed in 1982. It is often the regulatory mechanisms that impose such a transition to a new system. It will be noticed that in this case the system migrates gradually, loses the benefits of the previous model (a higher level of adaptation and inclusion of situations that are considered to be manageable within that profession), but gains the advantages of the new model (mainly in terms of safety).

The benefits of each of these models can be assessed on the basis of different beliefs and different approaches.

One might think that applying the resilient model to deep-sea fishing skippers does not represent a major problem, since their accidents have no major consequences outside their own profession. In the end, this is a choice that must be respected. On the other hand, the use of the resilient model in medicine raises complex ethical questions in relation to these two contradictory approaches: providing access, hope and care to everyone in all circumstances (which the resilient model does better than the ultra-safe model) and at the same time doing nothing that could harm or injure the patient (first do no harm) (which the ultra-safe model or even the HRO model do much better than the resilient model).

One may also consider that the nuclear model is not safe enough and call for even more standards and protocols to be introduced to cover situations that are considered to be more and more improbable. This is being done, for example, after Fukushima; having tested all the power stations in the entire world for the risk resulting from voluntary crashing of a passenger aircraft after the terrorist attack on the Twin Towers on 11 September 2001, the same power stations now have to be tested for their seismic risk and their risks of flooding and dedicated reinforcement measures must be taken. To some extent it will be necessary to bring what had been thought impossible into the realm of the possible and apply the demands of ultra-safe systems to this new possible situation. This strategy, which is typical of ultra-safe systems, is the exact opposite of a resilient solution: it reinforces the idea that the system is able to defend itself properly against known risks, but by doing this one refuses to learn to improvise to deal with a new exceptional surprise which will certainly occur one day (tomorrow? in 5, 10 or 20 years?) in one of the 500 or so nuclear power stations that are in operation throughout the world.

But would it be realistic for the nuclear industry to adopt a different strategy? A truly resilient strategy? Imagine that following Fukushima, the global nuclear industry decided to train its operators to give them greater resilience and recommended training them to deal with unexpected situations,[19] including conditions that have never been seen. It would then no doubt be necessary to train teams of operators to improvise and depart from procedures. It would also be necessary to be coherent and accept that there must be a two-speed system, one ultra-procedural one similar to the one that exists today, which maintains the existing exceptionally high level of safety ($1*10^{-6}$), thanks to strict compliance with procedures (a safety style which applies to 10^6 working hours[20] i.e. 45 years) and the other based on training a few expert operators, who are present in each power station, duplicating the standard operators, who are capable of improvisation, who would be deployed once in every two generations, or three times a century… and who would have to be banned from accessing the controls during the 27 years when no exceptional surprises arise, to avoid dangerous, unnecessary improvisation. Once the terms of this equation are set out, the answer is already clear: it is not feasible.

Another current example: many people consider that the financial crisis involving sub-prime mortgages and European debt that has afflicted the world since 2009 had the same roots: an excessively resilient model, driven by a minority of actors and in which profits are maximised by taking unreasonable risks, while intentionally masking and complicating transactions and financial products to

[19] The unexpected being referred to here is not the surprise of a logged breakdown occurring for which a procedure exists. Of course breakdowns and problems are not reported in advance, but they form an integral part of the ultra-safe model, with operators who are trained to respond. We are talking here about situations that have never been encountered before and for which no written procedures exist. It would therefore be necessary to improvise.

[20] This safety level of 10^6 is the guaranteed level for risk analyses at the design stage for the aviation and nuclear industries.

thwart all the supervisory procedures. In short, many ordinary people now consider that this system should change its approach and adopt more circumscribed, procedural safety rules, and depending on the political convictions of those involved, this would result in it at least adopting the rules of an HRO system, and in some cases becoming a totally supervised ultra-safe type system in which the actors involved have no autonomy at all. The fundamental nature of this system, however, is precisely its capacity to generate money competitively (freely) in order to finance the market; in short, this requires conditions that run completely counter to those of a controlled system.

It is no surprise that the proposals to improve moral standards in the global market economy and the role of the banks, to regulate and circumscribe their activities and even to (re-)nationalise them in the most extreme cases, although it was regularly discussed at G7 and then G10 summits, never went beyond the discussion of good intentions. The reason for this is simple: the models that would have to be called into question are above all societal models, with values and beliefs built on the successes of the past. No-one is really ready to abandon these for the sake of a hypothetical improvement in safety (whose results would not be seen until the entire process of transformation had taken place, and which would give rise to difficulties experienced immediately).

It is not possible to impose a completely new model of safety against the will of local actors and contrary to values that are considered essential to this system.

These underlying values must be shifted first, before making any claim to make people adopt a different safety model.

The lesson from this is simple: changing the safety model means changing the system. If the conditions are not met, and sometimes it is necessary to accept this fact, it is no good tilting at windmills or inventing solutions that have no chance of success.

The properties of three models of safety: the resilience model, the HRO model and the ultra-safe system model.

	Resilience model	HRO model High Reliability Systems	Ultra-safe model
Examples	Himalayan mountaineering professional fishing combat aviation international finance hospital emergencies	Merchant navy, Air freight, Naval armed forces Fire service oil industry Operating theatre	Civil aviation, Trains and metro services, Nuclear industry, Medical biology, Radiotherapy
Rationale	Taking risks is the essence of the profession. People take risks in order to survive in the profession.	Risk is not sought out, but it is inherent in the profession.	Risk is excluded as far as possible.
Key cultural trait	Fighter spirit, cult of champions and heroes.	Cult of group intelligence and adaptation to changing situations.	Cult of applying procedures and safety organised by an effective supervisory organisation.
Characteristics of accidents and lessons learned in terms of safety	Frequent, occur in various places, not many victims at a time, little media interest. Do not affect the profession. Only successes are really analysed (hard luck for the losers); People learn through adversity by analysing the stories of heroes who have survived in exceptional conditions.	Frequent, variable numbers of victims, media pressure sometimes considerable but accidents always result in enquiries and analyses. Learning from past failures mainly involves managing the same suboptimal situation in future (improving detection and recovery, actions focused on managing consequences).	Rare, large numbers of victims and extensive collateral damage, powerful media pressure. What is learned from past failures is essentially never to allow exposure to similar conditions again (better prevention, actions focused on eliminating causes).

(continued)

(continued)

	Resilience model	HRO model High Reliability Systems	Ultra-safe model
Principal feature of the model Who creates the safety	The expert skills of actors and their accumulated experience in the technical domain.	The ability of the group to organise itself (roles), to provide mutual protection to its members, to apply procedures, to be suspicious of excessively routine simplification of the situation, to adapt, perceive changes in the context and make sense of it.	The ability of the regulators of the system to avoid exposing front-line actors to unnecessary risks. The ability of front-line actors to follow protocols.
Standard types of operator training to improve safety	Learning through shadowing, acquiring professional experience, “training for zebra”, working on knowing one’s own limitations.	Training in teamwork to gain knowledge of the capacity of the group and adaptability in terms of applying procedures to suit the context.	Training in teamwork to apply procedures and apportion the work even if abnormal events occur.
Objective safety level of these systems	Between 10-2 and 10-4	Between 10-4 and 10-6	Better than 10-6
Capacity for progress within the model	Factor of 10	Factor of 10	Factor of 10

Each model represents a response to a type of environment, has its own rules for optimisation and offers little scope for mixing with any other model. It will be noted that safety can be improved within of these models by a factor of 10; a mixed form that incorporates some features of one safety system and other features of another will generally offer no results in terms of improvements in safety (it will be unsuitable or even counterproductive).

A Few Additional Rules When it Comes to Taking Action

We have just seen that building safety is not simple. It is necessary to recognise the risks, to build defences and above all to be realistic and choose the correct safety model. That, however, is not enough. Safety conceived in this way, however relevant it may be, needs to be introduced and maintained in a way that always stays in touch with those working on the ground. The management has a vital role to play in this, both in influencing behaviour (not just directing it) and in understanding clearly what the safety plan does not cover. Finally, it is also necessary to win the battle of the future and to avoid simply designing a system that corrects the errors of a past that no longer exists.

The role of management: to do well what it has been decided to do, and to have a clear knowledge of what it has been decided not to do

It makes no sense to create a safety strategy, however good it may be, unless it is understood and communicated. The management needs to have understood the aims (what must be achieved) and also the impasse situations that have been accepted and the reasons for these (what it has been decided not to do, as a trade-off for other commercial or business benefits that one wishes to retain). Education and training of middle management[21] and front-line managers in these two challenges are fundamental to a successful process. There is a plethora of studies in the literature that address this area [56, 57]. A very good summary is presented in the industrial safety manuals of the Institut pour la culture de sécurité Industrielle; these represent an important source of inspiration for this section.

Doing well what it has been decided to do: *the key role of middle management*

The traditional role of a manager is to manage, which means to carry out his duties as well as possible, to plan the activities or directing others. One should also add: a willingness to influence and offer guidance or orientation for his workers. This is the skill that makes a manager a leader. It is vital in order to make improvements in safety, since this is another area where collective mobilisation absolutely requires the management to show leadership, defined as the manager's ability to influence behaviour so that it becomes safer.

On the one hand, each person manages his own priorities in accordance with his own working context and the messages that he receives. People are generally attentive to the concerns of their own line managers, even if they are not explicitly held to account: in other words, if the line manager is not interested in a particular area, it is not very likely that his workers will be interested in it

[21] Middle managers are managers who work at an intermediate level in the hierarchy, between the executive level and front-line managers; they typically run a functional unit, source Uyterhoeven [54], Thakur [55].

either! On the other hand, the fact that personal safety naturally involves every individual, his integrity and his health does not necessarily result in spontaneous mobilisation in this area. For that to happen, each person needs to be aware of the challenges, convinced of the aims and their actions need to be coordinated. In brief, the management's own behaviour in relation to safety constitutes a message that carries much more weight than all the slogans posted on the company premises. These things demonstrate the true value of safety to the enterprise and they strongly influence the extent to which employees are motivated to behave safely.

The management has a key role in translating and monitoring the safety plan. This skill can be defined according to seven fundamental principles which are described in the box below.

BOX: Seven principles for leadership in industrial safety (source: ICSI, 2011)[22]

Principle 1 **Create the safety vision (which must be coherent with the management's values and principles)**

- Embrace and promulgate the safety policy of the enterprise
- Give safety the ranking that it deserves in relation to other challenges
- Imagine the future situation that one wants to see, based on the diagnosis of safety
- Set targets that are specific, measurable and achievable
- Build the vision collectively
- On the basis of the vision, define the principles of responsibility and the expectations in terms of behaviour

Principle 2 **Give safety the place it deserves in the organisation and in the management, and guide it in everyday practice**

- Integrate safety at every level in the organisation
- Clarify everyone's roles and responsibilities
- Define a progress plan that sets out the vision
- Systematically catalogue the obstacles that exist
- Make sure appropriate resources are available

Principle 3 **Ensure the safety vision is shared: influence, persuade and promote information feedback**

- Regularly reiterate the aims and expectations in terms of behaviour
- Reiterate these messages
- Communicate clearly
- Organise and promote the observation and classification of situations involving risk, including detection of weak signals
- Create a climate of trust and promote transparency
- Encourage the emergence of good practice; encourage and guide initiatives
- Remind people that safety is everyone's business

(continued)

[22] ICSI, Leadership en sécurité, Cahiers de la sécurité industrielle, accessed on 29 December at http://www.icsi-eu.org/francais/dev_cs/cahiers/CSI-LIS-pratiques-industrielles.pdf.

(continued)

Principle 4	**Be credible: exemplary behaviour and coherence** • Ensure that all actors are sufficiently competent to allow them to take the safety aims on board • Be competent, fair and consistent in judgements on safety • Be an example in terms of compliance with safety requirements and commitments, even in situations where things have gone wrong • Get personally involved in rolling out the Action Plan • Safety • Be capable of questioning yourself and questioning attitudes, including those of your superiors • Justify your decisions
Principle 5	**Promote team spirit and cooperation across the organisation** • Develop exchanges of views to resolve safety problems • Ensure that coordination resources exist to allow an overview of risks • Promote sharing of tools and methodologies • Bring safety officials together with operational employees on the ground • Make sure everyone feels integrated and has a sense of solidarity • Create connections where aims appear contradictory • Ensure that the traditional practices of the group are not in conflict with transparency or collective progress
Principle 6	**Maintain a presence on the ground to observe, listen and communicate effectively** • Organise visits on the ground • Organise regular meetings with the various professionals • Involve service provider enterprises in site visits, encourage and promote front-line access for managers of service provider enterprises • Emphasise what is going well and reiterate the lessons learned from past accidents • Catalogue problems that people experience in carrying out instructions and look for solutions together • Provide feedback to the actors involved about observations on the ground • Meet the victims of accidents
Principle 7	**Recognise good practices and apply sanctions justly** • Highlight good safety initiatives • Choose key moments for rewards and raising awareness • Communicate non-emotively about what is unacceptable and the related rules on penalties (possibly on a sliding scale) • Carefully analyse the context (technical and organisational environment, training) before applying any penalties and take care always to be fair and just • Be able to justify the penalty completely transparently

Have a good understanding of the safety trade-offs that are accepted in the action plan that is chosen

The management should also understand the measures that have been removed from the safety plan and the reasons for making these sacrifices (often strategic, financial or trade-offs against other priorities).

These decisions make safety more fragile and they need to be taken into account according to a dedicated strategy which involves reducing exposure, where possible, to those risks for which neither procedures nor training exist, and discussed at meetings to the point where operators are at least capable of detecting the situations, as well as providing keys to avoiding them.

For example, modern automated aircraft are protected from stalling by electronic safeguards that take back control from the crew when the flight characteristics breach threshold tolerance values (due to the airspeed or unusual attitude). Under very rare conditions, however, these aircraft may lose these electronic safeguards and the crew may face a loss of control and a stall. The aviation industry has chosen not to fully train crews to deal with these exceptional cases (for reasons of time, cost and technology—simulators are unable to mimic these conditions). In this case it is necessary for everyone to be fully aware of this gap in their competence and to learn to minimise exposure to these exceptional conditions, either through strategic anticipation to avoid conditions in which such stalls can occur (particularly a total loss of airspeed information), or by responding immediately when the alarm signals that these conditions are approaching.

Another example: most health care institutions have risk maps that do not include the risk of care being incorrectly managed by juniors in training posts when they are left alone on call in the hospital on the wards or in emergency departments, particularly at night, on public holidays, in August or at weekends. Hospitals do usually integrate the risk of errors by juniors in their maps, but they do this by stating that protection exists due to their supervision by seniors. In this situation, however, there are no seniors. The hospitals do not include this risk in their analysis because it would force them to avow an attitude which is virtually illegal and unjustifiable, since according to the official protocols it is unthinkable to leave trainees alone [58–60]. In these cases, executives should be aware that no protection exists against this risk and should organise the work as best they can to take this exposure into account: clearer instructions for calling seniors, words actually used when discussing this subject with juniors, teaching generic measures to safeguard patients etc.

Final example: the official risk map for professional sea fishing skippers on trawlers in the North Sea places considerable emphasis on the risks of collisions with tankers or ferries [6, 56]. The resulting response to their safety plan involves protecting authorised fishing zones by keeping them separate and avoiding overlaps with major shipping routes and Motorways of the Sea, and involves fitting these vessels with systems to detect other vessels (ARPA-Automatic-Radar Plotting Aid). This risk map, however, fails to take into account two factors relating to the fundamental economics of a highly competitive profession which is subject to quotas (first come, first served, best paid): “the fish are not subscribers” to the zones reserved for fishing, and the anti-collision systems broadcasting your own position (as intended) may therefore attract other fishing skippers to your fishing ground. It is not surprising that fishing skippers often work outside the permitted zones and intentionally turn off their anti-collision systems so that they cannot be tracked, but this situation is not covered by the risk analysis. Once again it is important to ensure that these practices are open to discussion among skippers (rather than leaving them under a “code of silence”) to ensure that the risk matrix being used is not based on naïve assumptions.

These three examples set out quite well the four commonest reasons for excluding an (identified) risk from the safety plan: excessive interference with the business model, too rare, too expensive to manage, too inappropriate to be admitted.

Practical rule: it is always useful, at the end of the process of building a safety plan, to go through all the areas that have been sacrificed, using the above categories to catalogue and classify them. The risks that are identified as a result should be treated in a specific way, which is more about sharing information and awareness about the fact that they exist, that they must be identifiable and that every effort must be made to avoid encountering them.

Thinking in the Future Rather Than the Past

Risk management techniques are essentially built using the rear-view mirror. They read the past to generate warnings for the future, and they aim to stabilise the world to safeguard the usefulness of lessons drawn from the mistakes of the past. The process is integrated in the form of an evolution rather than a revolution, and the inferences in terms of risks tend to be linear.

Unfortunately, the world is not quite so linear: it evolves through ruptures and sudden alterations, sometimes after two or three decades of stability. It is technical innovations that most commonly lead to professions being suddenly displaced. In a very short time, a technological system can become completely obsolete, together with its safety rules. In less than 15 years, for example, this change has occurred with the replacement of silver-based photography with digital photography, the end of cathode ray tube televisions and the explosion of mobile technologies. These revolutions have also affected a number of major industrial systems: the ongoing transition from oil prospecting and operations to large-scale oil shale recovery, the ongoing transition from essentially human-based air traffic control to automated regulation (datalink), the arrival of new and lighter building technologies and materials permitting structures to be built twice as quickly with characteristics that had never been achieved before in terms of height and on fragile ground, the arrival of new automobile engines, the switch from invasive surgery to non-invasive surgery by percutaneous or natural routes, the expected end of blood transfusions, as it is replaced by the production of blood from stem cell cultures, etc.

The essential point to remember is that the technology never changes by itself. It also changes the way the system is organised, its business model and its actors (new arrivals benefit from the technological leap to replace the old ones, using the example of digital photography), in short the whole model is replaced, together with its entire risk map and the defences that need to be built.

As a result it is essential to keep one eye on the horizon and constantly question the safety model that has been built on the foundations of the past. As technologies move forward more quickly, prospective methods may prove to be more effective than retrospective methods at avoiding the accidents of tomorrow.

Safety thinking in a system undergoing rapid change: the example of medicine. Medicine has entered into a period of rapid change, and a major crisis is coming in the long term, at least over the next 10 years (by 2020 or 2025). This situation is true in many industrial and service sectors: whether these are large sectors such as the nuclear industry in the post-Fukushima age, international banking and finance and the search for a new model, the aviation industry which is undergoing global restructuring, the oil industry which is facing the exhaustion of natural resources… or smaller high-risk sectors such as professional fishing, whose survival is threatened on a daily basis.

There are four types of forces that are simultaneously acting on all these sectors:

- the advent of radical innovations affecting both the substance and organisation of the work: in medicine: minimally invasive surgery by natural routes, genomic and personalised medicine, plus a series of other discoveries (oral chemotherapy etc.) which are spectacularly reducing the length of stays in hospital and thus creating the need for a different model of a short-term hospital (fewer technical hospitals, fewer beds, supplemented by a "hospital at home"). The equivalent in industry could be the change in the traditional oil exploration industry to intensive exploitation of oil shales…
- a sociological transformation of what is on offer and the professions involved; a drastic reduction in the number of surgeons, in favour of interventional professionals using minimally invasive surgical techniques (cardiologists, radiologists, gastroenterologists, ophthalmologists) resulting in a challenge to the historical position of operating theatres and the possible relocation of some of these procedures away from hospitals and into local primary care practices. We are also seeing a huge feminisation of health care actors and doctors in primary care, with a tendency to create joint practices in small towns, where everyone works part time and as a result rural areas become complete deserts with no medical provision at all. This leads to the introduction of telemedicine and the delegation of medical work to local nursing staff, and then from nursing staff to patients and their carers at home. In industry, one example of the equivalent change is the growing arrival of professions related to new forms of energy, whether it is solar, wind or fuel cells for the forms of transport of the future (to what extent will we still need engineers specialising in traditional combustion engines by around 2025?);
- an unlimited demand for safety with powerful pressures in the direction of transparency and external supervision;
- and clearly an unprecedented financial crisis, which is bringing the safety model face to face with the economic model more than ever before.

And Where is the Safety Culture in All This?

The reader of a book on risk management and safety management will expect such a book to address the safety culture and may even expect it to be quite a central subject. You will have noticed that this one does not. The aim is not to deny or reject the value of this concept, but to give the concept its true value in terms of the scale of its contribution towards safety. The first thing that is noticeable when doing this is the huge variability in the way the concept of the safety culture is used in the literature and the meanings that are given to it, so that one might conclude that the culture is closely linked to the safety model, but is rarely a concept that permits direct, primary action to improve safety. The time required to bring about changes is long, very long, and the process of enhancing the culture requires real perseverance in order to reap the rewards.

The themes of the safety culture and the safety climate are among the most popular subjects for publications in scientific journals specialising in industrial risks (and risks in public services, transportation and medicine). Most of the articles and books propose tools to evaluate the culture, especially questionnaires.

What, however, can truly be learned from these concepts in order to improve the safety of the system? This question deserves to be asked, because the answer is rather uncertain.

Cultures and Climates (of Change, Effectiveness and Safety), Multiple Areas of Ambiguity and Confusion

There are seven characteristics that dominate the literature on cultures; almost all of them raise fundamental questions about the usefulness of this concept for the purpose of improving safety.

1. Cultures are about values (significant ideas) and norms (expected behaviours) which are (1) moral, shared by all the individuals in a given community (social mores, relationships between men and women, relationships with the truth), (2) ethical (unacceptable conditions for success or failure in the community), and (3) social (definition of success, hierarchical distances, relationship with uncertainty, roles and expertise).
2. The concept of culture was first used to characterise national communities or enterprises long before it was extended to the specific context of the safety culture. In the context of national cultures, the work of psychosociologist Geert Hofstede [61] is a well-known reference point. It identifies five dimensions, and by combining these national cultures can be classified in relation to each other (without making value-judgments). The five dimensions are (1) the degree of hierarchical distance, (2) the need to reduce uncertainty, the degree of tolerance that a culture is able to accept in the face of concerns over future events (3) individualism versus collectivism, (4) the masculine macho versus feminine

dimension and (5) orientation towards the short or long term; links to traditions if the orientation is short-term, values of economy and perseverance if the orientation is long-term. Other major contributions have been put forward to characterise enterprise cultures (or organisation cultures or corporate cultures), particularly by O'Reilly, Chatman and Caldwell [62] who identify seven dimensions: innovation, stability, respect for people, focus on results, attention to detail, group solidarity, competitiveness and desire to win (aggressiveness). Finally, it would be difficult not to quote the important contribution made by sociologist Edgar Schein [63] who identifies three levels in a business culture: a visible level (artifacts) which shows the observable behaviours and rituals (this is typically the level that reflects the concept of the climate of an enterprise), the level of conscious values (values), which comprises the shared beliefs about the enterprise, its strong points, its weak points, its enemies, its friends, and finally a third level (the organisation's tacit assumptions) consisting of the tacit values, unconscious aspects or taboos that are shared but must not be named by the actors (unspoken rules), for example: "in this hospital we practice euthanasia for patients reaching the end of their lives in order to control the workload for staff".

At the outset none of these approaches explicitly referred to a value-classification of nations or enterprises, but they were all quickly used by other authors to describe "good cultures" and "bad cultures". From this time all kinds of difficulties began to arise.

Clearly the first problem in seeking to classify one culture in relation to another is to state what result is desired: in fact nations or enterprises can be classified according to their commercial performance, their capacity for change, their safety and many other criteria. It is no surprise that the classification systems in relation to a "good culture" will therefore differ depending on which criterion is used. Worse still, a good culture according to one scale (capacity for change, or efficiency), may turn out to be a culture that performs less well according to another scale (for example safety). So the first level of difficulty is: a culture is never good according to all the advantages that one might value according to every dimension. If enterprises choose to talk in terms of a "good culture" in the area of safety, this may lead them to adopt cultural traits that may be unhelpful or may even handicap them in relation to other key aspects of the challenges that they face in order to survive.

3. The concept of a safety culture is not homogeneous in any "genetic" sense. The same term is used to refer to very different theoretical approaches.

- Helmreich [64], Flin [65], Guldenmund [66] and many other authors (most of them following the other approaches set out below) have addressed the subject of the safety culture by looking through the prism of psychosociological theories on small groups and the roles of leaders and front-line managers, while prioritising the way in which front-line operators view their working environment. Many questionnaires have been developed on this basis, both on assessing cultures and on assessing the safety climate. It is no doubt due to these questionnaires, which are readily available, that these approaches have become so popular in the industrial world when "diagnosing the safety culture and diagnosing human and organisational factors".[23] The points that are identified in relation to a good safety culture by such "questionnaire diagnostics" are: a democratic leadership style, respect for everyone's role (in the hierarchy), respect for procedures, absence of a blame culture, ability to report errors/events/incidents without punishment, a sense that the hierarchy is listening, a high level of solidarity and mutual assistance within the group, low numbers of industrial accidents etc.

- Other authors have focused their definitions of a good safety culture on the way in which the management (middle and top management) deals with incidents and accidents (Westrum [67], Reason [68]) by insisting on the need for an in-depth analysis; some have gone even further by insisting on the sanctions that should be linked to these undesirable events, while pointing out the absolute need to maintain the system's ability to avoid judicial consequences when human error is involved—since this is necessarily involuntary—(concept of just culture [69, 70]).
- A very wide-ranging theoretical framework of organisational theories has provided the inspiration for approaches that address governance cultures and the macro-scale organisation of the system, and ultimately these are quite distant from perspectives centred on small groups and operators. This work has fed into the development of knowledge on enterprise cultures, for example by linking the quality of production, the capacity for innovation (climate of creativity [71]) and different families of cultures: tribal cultures, change cultures, cultures reliant on hierarchies and rational cultures[24] and by evaluating the specific characteristics and the progress that can be achieved in each type of culture.
- Still in the context of organisational theories centered on risk, the HRO (High Reliability Organizations) approach should definitely be mentioned. This specifically describes a good safety culture as consisting primarily of the ability of the group to adapt to non-standard situations, stressing the importance of leadership, expertise and everyone playing their role, and above all resilience or even improvisation, two ideas that were little discussed (or were even contradicted) in earlier streams. The HRO diagnosis of a culture is carried out not by using questionnaires but rather by auditing organisations.
- Ultimately, others have very strongly (exclusively?) equated the safety culture with the quality culture, from the perspective of improving the productivity and performance of the system; examples of this are the Toyota way [73] and Lean management. Once again, we are quite a long way from the earlier theories, with a culture that prioritises an organisation centred on the flow, gives the front-line management a key role in reducing the errors that cause falling performance, and manages quality in the production line, while considering the issue of serious accidents only to a very limited extent.

[23] Daniellou F, Simart M, Boissière I. Human and organizational factors of safety: state of the art, ICSI, http://www.foncsi.org/media/PDF/CSI-HOFS.pdf.

[24] A good summary of this whole approach in Braithwaite et al. [72].

This list is a long way from reflecting all the contributions and theoretical currents that exist in the area of the safety culture: different theories, different messages and a different focus. Most of the time, occasional readers and users cannot be aware of this range in its entirety and find themselves locked into a single point of view or approach. As a result they do not understand the contradictions that may arise if they mix an approach centred on the production line around a Toyota Way or Lean Management type culture, while simultaneously stating that they aim to be an HRO organisation and at the same time maintaining—in different circles—that the priority of the enterprise is to develop the culture and adopt a climate that promotes change so that it can address the coming challenges posed by new socioeconomic conditions.

In short, the phrase "changing the saefty culture" can easily hide major ambiguities and can lead to huge disappointment if it is actually rolled out at the operator level with no precautions. In many cases, fortunately—or perhaps unfortunately—the use of this terminology is no more than a fine-sounding form of words intended for external consumption which has no major consequences internally and no real usefulness. Having said that, every enterprise does have a culture, and it is perhaps more important when intervening in safety to understand all the contradictory aspects of that culture.

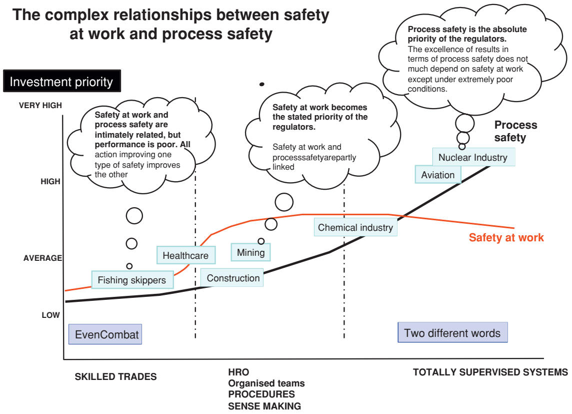
Cultures and accident rates in civil aviation. Helmreich [74] showed, building on the work done by Hofstede in the early 2000s, that crews from highly collectivist countries in which considerable hierarchical distances are maintained (Central America, Central Asia) had accident rates in passenger aircraft which are two and a half times higher than those in Western countries, which are characterised by more individualistic cultures without any hierarchical distance. Initially the interpretation naturally tended towards the idea of making value-judgements when classifying cultures and towards putting forward the idea that everyone should adopt the characteristics of Western cultures since these are associated with the best safety results. This hypothesis, however, very soon ceased to be used in such explicit terms. Ethical grounds played an important part in this (the need to avoid insulting these countries or their national values), but the real reason was more trivial: the study had simply revealed something quite obvious: the design of a complex system (like an aircraft) is profoundly marked by culture; it is much easier for users who share this culture to use the product correctly: modern, automated aircraft require very direct collaboration between crew members and expect subordinates to be able to constantly express surprise and question the actions and decisions of their boss; countries that maintain more distant hierarchical relationships inevitably have difficulty working within this model. To some extent there is no such thing as a "good culture", but there are bad "marriages".

4. **A poorly understood link between two drivers that produce the safety culture: worker safety and site and product safety.** Quite strangely, the literature has developed two parallel frameworks in which safety culture diagnostics are applied within enterprises: safety at work and site and product safety. Three observations can be made about this ambiguity:

– the priority given to each of these two areas depends on the maturity and the public priorities of regulators in various types of industries and public services. The two priority curves intersect. For the most immature and least safe activities (skilled trades, medicine), the public priority mostly concerns production safety (for example patient safety); in industries which are more labour-intensive and more mature than skilled trades and have powerful regulators, a higher priority is given to safety at work (reducing the number of industrial accidents). Paradoxically, for the safest industries (in terms of industrial accidents), the priority switches back to process safety; these industries, which are the safest (in terms of the risk of accidents) often have no more than average performance in terms of safety at work (and certainly lower than the industries in the previous group);

– other than general ideas, we are not very aware of the theoretical links that exist between these two areas of safety. The link is clearly a complex one, since excellence in one of these two areas is rarely associated with excellence in the other;

– this ambiguity continues in the use of many tools to measure cultures and safety climates, particularly questionnaires, which have often been validated for only one of these two areas but are still used without any precautions to assess the other.



5. **It is possible to change a safety climate quickly, but a safety culture cannot be changed quickly.** The concept of a safety climate, which was first inspired by Schein (op. cit.) refers to objective aspects (facts), while the concept of the safety culture refers to subjective aspects (values). It may

be possible to change certain elements of the safety climate quite quickly through dedicated actions by the management or organisation, but the values that characterise a culture cannot be changed as quickly. Having said that, it is possible to significantly influence the culture by making authoritative changes to the fundamentals of the technical system and introducing major changes to the economics of the system, but of course this is beyond the limitations of an ad-hoc intervention in an enterprise in an industrial sector or service (such as a hospital or a bank). In short, the market economy dictates the culture rather than vice versa. The levers for change are systemic, not local.

Changing the culture in civil aviation: a systemic lever far ahead of the human factors lever. The passenger aviation sector was long the domain of heroes, in which captains were in command under God alone, deciding which route to take and making exceptions to procedures whenever they considered it appropriate. The introduction of air traffic control after the Second World War represented the first limitation on this autonomy, but it was above all the tremendous global standardisation of the supervision of the flight system, the arrival of automated aircraft that erased handling differences between pilots and the recording of all actions taken by the crew in the cockpit (systematic flight analysis) that definitively tipped the culture of civil aviation in the 1980s towards the ultra-safe model. The introduction of crew training during the 1990s and the initiatives towards voluntary no-blame reporting, which was given a high profile in the media, accompanied these changes rather than being the real reason for changes in the culture (which is now characterised by equality among actors, a high level of transparency surrounding incidents and teamwork centred on coordination and monitoring of procedures).

6. **There is no ideal culture, but there are cultures that are suitable for every situation.** This perspective has gradually become established as the only one that can cope with the real situation. All normative approaches to this area have turned out to be counterproductive. We have seen in the paragraphs above that there are several different models of safety rather than just one. It makes sense that these different models of safety, which reflect different trade-offs between flexibility, competitiveness, adaptability and safety performance, should be based on different ways of managing the safety culture.
7. **The development of the characteristic values of a culture takes a very long time.** Some people speak of a generational lever. None of the standard risk matrices or safety action plans cover such long periods of time.

In the end, assessing the safety culture of a production unit is a useful activity and forms an integral part of a diagnostic process. It requires quite an in-depth knowledge of the theories behind the measurement tools, to avoid the occurrence of contradictory effects. It is not, however, sufficient in itself. The interpretation of such assessments is always relative, because it depends on local challenges (which need to be properly analysed and understood). Finally, this measure makes it possible to ascertain the margin available for progress in the domain of safety that is available to the enterprise.

> If a local safety intervention has to be undertaken in an enterprise within a specific period of time, rather than expecting to change its culture, the opposite approach should be taken: deducing (from an assessment of the culture) what margin exists for real progress to be achieved by the enterprise, in view of its culture.

This diagnostic process will promote the identification of the safety model that best characterises the environment (and the needs) within the enterprise being audited. In short, the culture of an enterprise cannot be changed by a single intervention motivated by a demand for safety. There is no action at all that will achieve this. It is possible, however, to understand and identify the culture that does exist, in order to assess what margin exists within that culture to improve the results (reverse approach).

If a more ambitious approach is adopted that claims to be able to change the culture within a profession, one must have systemic levers, change the demands of the business model at the level of the whole profession, take action at least at the regional level if not nationally or internationally and sustain this action over the long term (long-term intervention is vital, with regulatory systems designed for the purpose).

To the extent that the safety culture is a consequence of the economics of the profession and its safety model rather than a cause of that model, it is legitimate that this paragraph should come at the end of this text rather than at the beginning.

References

1. Reason J (1990) Human error. Cambridge University Press, Cambridge
2. Reason J (1997) Managing the risks of organizational accidents. Ashgate, Aldershot
3. Heinrich HW (1931) Industrial accident prevention. McGraw-Hill, New York
4. Hollnagel E, Woods D, Levison N (2006) Resilience engineering: concepts and precepts. Ashgate, Aldershot
5. Dekker S (2004) Ten questions about human error. A new view of human factors and system safety. Ashgate, Aldershot
6. Morel G, Amalberti R, Chauvin C (2008) Articulating the differences between safety and resilience: the decision-making of professional sea fishing skippers. Hum Factors 1:1–16

7. Roussel P, Moll MC, Guez P (2007) Étape 2: Identifier les risques a priori. Risques & Qualité en milieu de soins 4:239–247
8. Roussel P, Moll MC, Guez P (2008) Étape 3: Identifier les risques a posteriori. Risques & Qualité en milieu de soins V-1:46–58
9. Tengs T, Adams M, Pliskin J et al (1995) Five hundred life-saving interventions and their cost-effectiveness. Risk anal 15(3):369–390
10. Johnson C (2003) Failure in safety-critical systems: a handbook of accident and incident reporting. University of Glasgow Press, Glasgow
11. Reason J (1997) Managing the risks of organizational accidents. Ashgate, Aldershot
12. Marx D (2001) Patient Safety and the "Just Culture": a primer for health care executives. Columbia University, New York
13. Dekker S (2007) Just culture, balancing safety and accountability. Ashgate, Aldershot
14. Guldenmund F (2007) The use of questionnaires in safety culture research—an evaluation. Saf Sci 45:723–743
15. Guldenmund F (2000) The nature of safety culture: a review of theory and research. Saf Sci 34(1–3):215–257
16. Flin R (2007) Measuring safety culture in healthcare: a case for accurate diagnosis. Saf Sci 45:653–667
17. Lawton R, Parker D (2002) Barriers to incident reporting in a healthcare system. Qual Saf Health Care 11:15–18
18. Cullen D, Bates D, Small S et al (1995) The incident reporting system does not detect adverse drug events: a problem for quality improvement. Jt Comm J Qual Improv 1:541–548
19. Jha A, Hupeerman G, Teich J et al (1998) Identifying adverse drug events. JAMIA 5:305–314
20. Goldman RM, de Leval AP, Cohen MR et al (2004) Pitfalls of adverse event reporting in paediatric cardiac intensive care. Arch Dis Child 89:856–885
21. Vincent C, Stanhope N, Crowley-Murphy M (1999) Reasons for not reporting adverse incidents: an empirical study. J Eval Clin Pract 5:13
22. Evans SM, Berry JG, Smith BJ et al (2006) Attitudes and barriers to incident reporting: a collaborative hospital study. Qual Saf Health Care 15:39–43
23. Ricci M, Goldman AP, de Leval MR, Cohen GA, Devaney F, Carthey J (2004) Pitfalls of adverse event reporting in paediatric cardiac intensive care. Arch Dis Child 89:856–885
24. Dekker S (2007) Just culture, balancing safety and accountability. Ashgate, Aldershot
25. Classen D, Resar R, Griffin F et al (2011) Global trigger tool shows that adverse events in hospitals may be in times greater than previously measured. Health aff 30:581–589
26. Resar R, Rozich J, Classen D (2003) Methodology and rationale for the measurement of harm with trigger tools. Qual Saf Health Care 12:39–45
27. Rozich JD, Haraden CR, Resar RK (2003) Adverse drug event trigger tool: a practical methodology for measuring medication related harm. Qual Saf Health Care 12:194–200
28. Sjohania K, Sampson M, Ansari M et al (2007) How quickly do systematic reviews go out of date? A survival analysis. Ann Int Med 147:224–233
29. Hollnagel E (2004) Barriers and accident prevention. Ashgate, Aldershot
30. Gerstein M, Ellsberg M, Ellsberg D (2008) Flirting with disaster: why accidents are rarely accidental. Sterling Publishing, New York
31. Ostberg G (2009) Some intangibles in human handling of risks. Lund University, Sweden
32. Chateauraynaud F, Torny D (1999) Les sombres précurseurs: une sociologie pragmatique de l'alerte et du risque. EHESS, Paris
33. Amalberti R (2006) Optimum system safety and optimum system resilience: agonist or antagonist concepts? In: Hollnagel E, Woods D, Levison N (eds) Resilience engineering: concepts and precepts. Ashgate, Aldershot, pp 238–256
34. Woods DD (2005) Creating foresight: lessons for resilience from Columbia. In: Starbuck WH, Farjoun M (eds) Organization at the Limit: NASA and the Columbia Disaster. Blackwell, Hoboken, pp 289–308
35. Aslanides M, Valot C, Nyssen AS, Amalberti R (2007) Evolution of error and violation description in French air force accident reports: impacts of human factors education. Hum Factors Aerosp Saf 6:51–70

36. Amalberti R, Vincent C, Auroy Y, de Saint Maurice G (2006) Framework models of migrations and violations: a consumer guide. Qual Saf Healthc 15(suppl 1):i66–i71
37. Amalberti R (2001) La maîtrise des situations dynamiques. Psychologie Française 46–2:105–117
38. Rasmussen J (1997) Risk management in a dynamic society. Saf Sci 27:183–214
39. De Saint Maurice G, Auroy Y, Vincent C, Amalberti R (2010) The natural life span of a safety policy: violations and migration in anaesthesia. Qual Saf Health Care 19:327–331
40. Cooper D (2009) Behavioral safety interventions. Professional Safety: 37. http://www.behavioural-safety.com/articles/behavioral_safety_interventions_a_review_of_process_design_factors.pdf
41. Amalberti R, Auroy Y, Berwick D, Barach P (2005) Five system barriers to achieving ultrasafe health care. Ann Intern Med 142(9):756–764
42. Morel G, Chauvin C (2007) A socio-technical approach of risk management applied to collisions involving fishing vessels. Saf sci 44:599–619
43. Wilson R (1979) Analyzing the daily risks of life. Technol Rev 81:40–46
44. Amalberti R, Auroy Y, Berwick D, Barach P (2005) Five system barriers to achieving ultrasafe health care. Ann Intern Med 42:756–764
45. Perrow C (1984) Normal accidents: living with high-risk technologies. Basic Books, NY
46. Bainbridge L (1987) Ironies of automation. In: Rasmussen J, Duncan K, Leplat J (eds) New technology and human errors. Hoboken, Wiley publishing, pp 271–286
47. Woods DD, Hollnagel E (2006) Joint cognitive systems: patterns in cognitive systems engineering. Taylor & Francis, Boca Raton
48. Endsley MR, Garland DJ (2000) Situation awareness analysis and measurement. Lawrence Erlbaum, Mahwah
49. Endsley M (1995) Toward a theory of situation awareness in dynamic systems. Hum Factors 37:32–64
50. Sarter N, Woods D (1992) Pilot interaction with cockpit automation: operational experiences with the flight management system. Int J Aviat Psychol 2(4):303–321
51. Amalberti R, Deblon F (1992) Cognitive modelling of fighter aircraft's control process: a step towards intelligent onboard assistance system. Int J Man-Mach stud 36:639–671
52. Morel G, Amalberti R, Chauvin C (2009) How good micro/macro ergonomics may improve resilience, but not necessarily safety. Saf Sci 47:285–294
53. Grote G (2012) Safety management in different high-risk domains—all the same? Saf Sci 50(10):1983–1992
54. Uyterhoeven HE (1972) General managers in the middle. Harvard Bus Rev 50:75–85
55. Thakur M (1998) Involving middle managers in strategymaking. Long Range Plan 31:732–741
56. Hopkins A (2005) Safety, culture and risk, 1st edn. CCH Australia Ltd, Australia
57. Hopkins A (2007) Holding corporate leaders responsible. Keeping Good Co 59:340–344
58. Bell CM, Redelmeier DA (2001) Mortality among patients admitted to hospitals on weekends as compared with weekdays. N Engl J Med 345:663–668
59. Aylin P, Yunus A, Bottle A et al (2010) Weekend mortality for emergency admissions: a large multicentre study. Qual Saf Health Care 19:213–217
60. Young J, Ranji S, Wachter R et al (2011) July effect: impact of the academic year-end change over on patient outcomes. Ann Intern Med 155:309–315
61. Hofstede G (1983) Culture's consequences: international differences in work-related values. Adm Sci Q (Johnson Graduate School of Management, Cornell University) 28:625–629
62. O'Reilly C, Chatman A, Caldwell D (1991) People and organizational culture: a profile comparisons approach to assessing person-organization fit. Acad Manag J 34:487–516
63. Schein E (1985) Organizational culture and leadership. John Wiley & sons, Hoboken Ed 2010
64. Helmreich RL, Merritt AC (1998) Culture at work: national, organizational, and professional influences. Ashgate, Aldershot
65. Flin R, O'Connor P, Crichton M (2008) Safety at the sharp end: a guide to non-technical skills. Ashgate, Aldershot

66. Guldenmund F (2007) The use of questionnaires in safety culture research—an evaluation. Saf Sci 45:723–743
67. Westrum R (2004) A typology of organisational cultures. Qual Saf Health Care 13:22–27
68. Reason JT, Carthey J, de Leval MR (2001) Diagnosing vulnerable system syndrome: an essential prerequisite to effective risk management. Qual Health Care 10(suppl 2):i21–i25
69. Marx D (2001) Patient safety and the just culture, a primer for health care executives. MERS-TM. Columbia University, New York
70. Dekker S (2008) Just culture, balancing safety and accountability. Ashgate, Aldershot
71. Ekvall G (1991) The organizational culture of idea-management: a creative climate for the management of ideas. In: Henry J, Walker D (eds) Managing Innovation. Sage, London, pp 73–79
72. Braithwaite JJ, Hyde P, Pope C (2010) Culture and climate in health care organizations. Palgrave MacMillan, Basingstoke
73. Liker J (2003) The Toyota Way: 14 management principles from the world's greatest manufacturer, First edn. McGraw-Hill, New York
74. Helmreich R (1993) Attitudes towards automation across five cultures—NASA report/ University of Texas/FAA Aerospace Crew Research

Chapter 4
Human and Organisational Factors (HOFs): Significantly Growing Challenges

Abstract Although this book concentrates on the safety of complex systems, it seemed to me that it would be useful to finish with a presentation of the major historical discontinuities and the successive principles that have guided human and organisational factor (HOF) processes in enterprises.

Based on my view of the history of industrialisation, I can identify five important, successive stages in the management of HOFs. This realisation enriches and will inevitably modulate the strategy for systematic safety management.

The Productive Worker

The first step was taken in the late 19th/early 20th century, with increasing power of intensive industrialisation. It incorporated three fundamental concepts that were intended to "obtain more productive workers".

The first of these was Taylorism [1] which seeks to increase productivity by taking better account of (essentially physical) human characteristics and the demands of the work that has to be done. Knowledge of these two areas makes it possible to achieve the best compromise in terms of productivity in every working situation (mainly involving repetitive, physical work) to maximise efficiency without reducing or unnecessarily degrading the workers' abilities in the long term. The idea of a scientific way of organising work that can be analysed, taught and generalised, is indisputably the primary historical goal of HOFs.[1]

Taylorism was adapted but did not disappear, particularly in manual activities (it is not so easily applicable to intellectual activities by its nature). One in five

[1] Fordism was not very far from Taylorism and dates from the same period. It involved pushing standardisation to the extremes on the Ford production lines and compensating workers for this by paying them a high salary ("five dollars a day").

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology, DOI: 10.1007/978-94-007-6549-8_4, © The Author(s) 2013

workers still work under modern forms of Taylorism. The Toyota Way[2] emerged in the late 1970s and is merely a modern re-reading of the concepts put forward by Taylor. Quality control was placed at the centre of this system and personnel were better integrated, the process of working in rotating shifts was made universal, tasks were enriched (for example making them less monotonous on the production lines, with successive vehicles requiring different actions at the same position on the line for different finished products), and the line also allowed a little more time for personal adjustment (catch-up time etc.) while still working within collective time envelopes. Ultimately, however, the central idea was still rationalisation of the time taken to do the work. The sudden rise in the incidence of RSI (musculo-skeletal problems or Repetitive Strain Injuries) in the late 1990s [2] shows the extent to which some of the principles underlying Taylorism have already been forgotten, when these should have been retained as an integral part of our academic knowledge. No doubt the intense humanistic polemics of the 1980s went a long way towards the condemnation of Taylor, which was no doubt excessive, but it was the huge technical and methodological difficulty of transferring Taylorism to the complex intellectual work of our modern societies that finally killed off this approach in universities, and particularly in University courses in ergonomics.

The second step that was taken during this early period was definitely taking context and motivation into account, and specifically the psychological context of work. The work of Mayo [3], Gillespie [4] which which was carried out between 1927 and 1932 at the Hawthorne factory, laid the foundations for this approach. Mayo's first intervention was in an assembly plant at the Western Electric Company (which assembled radio equipment), with production lines predominantly employing female workers. It demonstrated the impact on productivity of addressing working conditions. He created an experimental workshop with better light and better working conditions. Production increased. Mayo returned to this factory a few months later, into the same workshop, and tested quite an astoundingly clever idea: in the experimental workshop, he reversed the improvements in working conditions that had previously been made. Production increased even more. He had proven that production is largely driven by the level of interest in workers, which is translated into additional motivation. This line of work was to inspire the sociology of work and it gave rise to many ramifications and developments which still have currency today.

The third step to be taken during this early period was the timing of selection. At a very early stage, selection by means of degrees or psychological tests looking for particular skills was found to be a third promising route, which formed a natural complement to training and organisation in reducing the natural variability of subjects and allowing better adaptation to the work, including their responses to poor conditions. The creation in 1916 by G. Stanley Hall, John Wallace Baird and Ludwig Reinhold Geissler of the Journal of Applied Psychology, which was

[2] The Toyota Way was invented by engineer Ohno on the Toyota production lines, and its aim can be expressed in terms of five zeroes: zero stock, zero faults, zero paper, zero breakdowns, zero delays.

devoted almost entirely to this subject of selection, is still a historical marker in the rise to prominence of the idea of selection.

These fundamental features of approaches to human and organisational factors underwent little further change until the 1960s. Knowledge of ergonomics and working conditions increased, but there was no revolutionary change.

Here is a summary of this early period.

- The emphasis was on operator productivity.
- The main type of know-how offered by HOFs during this stage was in the design of workplaces to allow greater productivity: analysis of the demands, knowledge of operator abilities and how to select for these, and solutions making it possible to improve working conditions.
- Safety was only taken into account as a variable supporting productivity (a well selected, well trained and well motivated worker has fewer health problems, produces more and causes fewer accidents).
- The scientific deployment of (analytical, contextual and selection) concepts was motivated primarily by the search for rapid and definite improvements in productivity.
- The management was not a target during this early period.

The Safe Factory

The second step accompanied **the increasing power of high-risk industries** in the late 1960s.

This step continued to be highly productive until the early 1990s, and expressed the overriding desire to **control the risks** that were known to be growing in industrial plants (mainly in the chemical and nuclear industries).

The accident at the factory in Seveso in Italy in 1976 when a chemical reactor overheated and caused a release of dioxin led to rapid acceleration in the need to create a science that would address both reliability and safety (Reliability engineering and system safety [5, 6]). Following the nuclear accident at Three-Mile Island in 1979 the community was finally convinced that this approach should be a priority and particular emphasis was placed on the human component of reliability, without altering the general principles of the approach.

The fundamental idea behind these approaches to calculating risk is that risk is equal to the frequency (of events) multiplied by the seriousness of their consequences. This fundamental equation still forms the basis for risk assessment methods (System-Risk Assessment or SRA) and by extension for methods used to assess human reliability Human-risk assessment (HRA) [7].

A large number of measurement and decision-making tools (to define what should be prioritised in a safety action plan) were developed in the 1980s; some versions of these became much more forward-looking at the design stage (for example HAZOP), while others added dimensions of the basic calculation process (FMECA calculates risk as frequency multipled by seriousness multiplied by detectability).

The concept of feedback emerged at this stage as an essential input for these calculation models.

Here is a summary of this phase.

- The context has changed in high-risk industries; the process used in the factory can create collateral damage far beyond the individual workplace.
- Safety has become the main focus of interest in the HOF process. Error analysis takes precedence over all other analyses of working conditions.
- Attention is still focused on the internal production tool.
- The analysis prioritises the process, its failings and the control procedures: there is a search for safe solutions to malfunctions, whether these are imagined or have occurred in the past, whether they are technical in origin (no procedure or incorrect procedure) or human (error), but the analysis is still technocentric.
- The analysis takes the form of a calculation and is supported by feedback.
- The risk has reached the point where safety is no longer a competing variable between operators. They all have to achieve excellence. The role of regulators and regulations becomes a major factor and applies to everyone.
- This phase, however, is still openly positive and there is an atmosphere of hope: there is a belief in excellent safety: it is possible to do better and it is possible to imagine better processes and better tools. The application of concepts of reliability inevitably leads to resolution of the problems that are identified. The path is a long one, but it will be followed and the accidents will disappear.

The Safe Product, Safety Challenges Raised by Design and Use

The third step accompanied the increasing importance **of public transport** in the late 1980s and 1990s: automated aircraft, TGV trains, guided metro systems etc.

For the first time, the analysis prioritised the quality of the commercial product (train or aircraft safety) rather than the site where it is designed and produced. The approach had become decentralised to encompass both design and use. Safety problems within the factory became secondary; problems relating to product use and product quality were the priority.

This decentralisation was both the product and the result of two revolutions:

- a technical revolution, with the large-scale introduction of information technology and electronics, making it possible to rethink the way the product is both designed and used in a modelled, supervised environment [8]. The result is a radical change in interfaces, extreme precision in guidance and flow performance (vertical and horizontal separation of aircraft) and a redefinition of the place of operators. A new scientific discipline is born: Cognitive Engineering [9]. Electronics has also introduced the possibility of an unbreakable safeguard to protect against errors by operators. Looking beyond the pilot's cockpit, the concept of network operation and network safety, which did not exist during the previous phase, becomes a central concern;

- a revolution in the allocation of roles and responsibilities. The design of the product influences customers well beyond the tool itself and requires usage rules that have an ever greater impact on the way their society is organised. For example, the automation of aircraft and air traffic control has profoundly changed the professional rules and the management rules within airlines (changes in career paths, training programmes and recruitment profiles). The more complex the system becomes, the more often its builders end up creating the entire operating system, together with its standards and the rules governing its deployment and use. These have a major influence on regulations, rather than simply being a selling-point for the tool itself. We saw in Chap. 3 that builders end up exporting their entire culture, including their vision of safety, the community and of the organisation of work within the hierarchical chain. It must be said that operators did not really have much choice: the offerings of a very small number of global builders were very similar and the real purchasers were the banks, which make decisions based on macroeconomic criteria and lease the products back to the operators.

Of course the designer is *ultimately* not the user. A system is put in place to permit sharing of the management of risks. The global distribution of products comes into conflict with different cultures, which have long sustained the debate on HOFs: the level of deviation that should be tolerated from the recommended use of the product as determined by the builder, the extent to which the operator can be allowed to appropriate and tailor the product designed by the agency etc. The introduction of Crew-resource management (CRM) in the aviation industry [10] to guide, train and improve safety in teamwork and the Swiss cheese model put forward by Reason (op. cit.), to cover the emerging need for system safety, symbolise the two most striking contributions made by this phase in terms of HOFs.

Here is a summary of this phase.

- The safety of the product that is delivered and the way it is used take precedence over production processes within the factory.
- On the technical side, the information technology revolution makes centralised automation and supervision of a network possible.
- Safety is based on product design, operation and use. New areas for development open up in terms of product use, particularly in terms of the safe organisation of teams (crew training and crew members) and technical organisation (suboptimal organisational choices and latent errors in the operating chain of command). The analysis still focuses on front-line actors and front-line management. HOFs do not yet have access to senior managers, but they are gaining access to middle managers.
- HOF know-how improves very rapidly during this stage; contributions are made on teamwork, team reliability, the safety culture and the reliability of the

hierarchical chain. The design process determines usage rules and carries out client modelling—including economic modelling, but HOFs are still quite separate from this highly important economic dimension, which is left to the discretion of major audit firms.

- The changes made during the previous phase (increasing power of regulators and reliability studies) of course remain in place.

The End of the Impossible Dream of Safety

The fourth step corresponds to the disappearance of a myth that everyone wanted to believe in: it was thought until then that applying science to safety interventions would make it possible to control the risk while meeting the demands of society and alleviate its concerns.

The disappointment came in the second half of the 1990s. The industry began to realise the irony of the tremendous efforts that were being devoted to safety. There would never be any escape from the pressures and growing investments in the area of safety. The momentum behind the demands from society was infinite and the demands were increasing, to the point that the efforts that the enterprise was making to invest in safety could never return to a break-even point. In fact the opposite was true: the more it improved, the more it had to invest in safety to the detriment of other investments in production.

The key characteristics of this phase have already been described in Chaps. 1 and 2.

For the first time the strong link was also becoming established between areas of safety that had been separate until that time: industrial safety on the one hand, safety of installations (particularly their vulnerability to terrorism) and environmental safety. The major disturbances in the climate and in the political world were feeding and increasing the risks taken in terms of factory localisation, the design and use of complex systems, which were mainly organised in the form of supervised networks. Risk map ping began to take place in a transverse structure in order to cover all these issues. Terrorism and threats (epidemics etc.) took on global proportions. Confidence in a future, which was being depicted in the form of linear, continuing improvement in safety, disappeared.

There are no limits to safety, and the safer things are made the greater the demands that are made and the heavier the penalties imposed on the remaining accidents. **Safety is not just a scientific question; it appears above all to be a question of social perception.**

At the same time, this loss of confidence in science was to open the way for the emergence of the **principle of precaution**[3] and approaches to victim compensation based on less and less scientific criteria.

[3] The principle of precaution was formulated for the first time in the Rio Declaration in 1992, and initially only applied to environmental risks. It was rapidly extended to the whole field of major medical risks. For example, read Gollier [11], or Guilhou [12].

We are seeing a process of distancing from sickness and death which is paradoxically increasing anxiety and heightening demands. Any premature injury (or threat) gives rise to social indignation.

This increase in awareness was to feed into the practice of modern industrial medicine, create opportunities for previously unknown levels of compensation, taking into account chronic disorders (the case of asbestos) which have the potential to shorten life expectancy and impair quality of life, in areas which are infinitely more subjective and complex to calculate than an immediate injury at the workplace. Safety was previously an obligation for industry, involving a commitment to deploy means; now it became a right for citizens, with a duty to achieve results, and it turned out to be susceptible to claims and monetary demands in which there was almost no limit to the levels of possible financial compensation. Equally paradoxically, the recent major improvements in convenience in both society and work (for example the 35 h working week in France) are giving rise to new and major difficulties in adaptation among workers, who are suffering from the lack of time, increasingly intense work and from a certain loss of humanity in their social relationships. HOF processes are intended to deal with suffering at work and suicide among professionals, who are treated infinitely better than their forebears, but their suffering is not alleviated by the history of those others or by the relative nature of their situation… it is completely personal and is based on their own experience of suffering.

The social cost of this form of safety is leading to one crisis after another in society: an insurance crisis, a competitiveness crisis, a liability crisis and in-depth debates on whether high-risk industries can continue to be located in the most sensitive countries at all. It also favours huge transfers of liability, particularly through the increased use of outsourcing.

Transparency, which is the most powerful concept underlying this “impossible safety” turns out to be terribly difficult to tame.

The societal system is entering a period of complete irrationality, to the point where it is resulting in the investment of greater industrial resources in know-how in relation to crisis management and management of economic collateral effects than on the prevention of accidents proper and the immediate damage that they can cause.

Here is a summary of this phase.

- The improvement in safety results in ever-greater demands for safety.
- The “last accident” costs infinitely more in terms of damage to the image and compensation than earlier accidents, which were the results of an industry that had not made any substantial effort.
- Despite, or perhaps because of the progress that has been made, workers are suffering more than in previous phases.
- There is a sudden move away from the idea that it is possible to achieve total control and reduction of risks and towards an acceptance of the idea of charting a course dynamically through the risks, with a need to achieve the best trade-off between risks in the various dimensions (accidents, economic aspects,

production, human resources, competitive strategy) to survive the inevitable crises that will occur (in one dimension or another).

- This phase is having a major impact on HOFs. It does not offer any really new models of safety, other than those of crisis management, but it is becoming necessary to consider using all the available models more systematically, deploying transverse approaches, taking a bird's eye view of situations, taking governance into account and above all offering advice which is no longer optimised in terms of a single dimension but in terms of the equilibria, trade-offs and sacrifices that provide the best returns. In other words, a modern HOF approach no longer exclusively answers the question that is asked, particularly in terms of safety; the process has to go all the way back to the overall management and reposition—or even moderate—the local response in a way that takes into account collateral effects on the other contradictory dimensions of the activity of the enterprise.

For example, patient safety (medical errors) is only one of the major risks that have to be managed by a medical institution. It is known that all actions taken in relation to that specific risk will have an impact on the other risks facing the institution and will often increase them. This in turn explains the fact that measures that are planned are ultimately rarely pushed through by clinical managers [managers also have to manage the economic risks (occupancy rates, billing, bad debts etc.) and resource risks (finding competent doctors and staff to remain open 24 h a day), the risks of non-conformity (fires, buildings, sites etc.) strategic risks (critical mass, political alliances, purchasing, shareholders etc.). A good intervention in terms of HOFs should answer the question that is asked while clearly identifying the collateral effects in terms of other risks, so as to come up with a decision whose effects and collateral costs are well understood and controlled by managers. Otherwise the intervention will be doomed to failure.

Uncertainty as a Future Risk: Future Risks as a Central Feature of the Present

We have now entered a **fifth phase for HOFs**.

This is characterised by a focus on future risks ("which do not yet exist and which we want to avoid") while all the previous phases were focused on past risks that people were seeking to reduce.

The major debate on environmental threats provides an excellent illustration of this new shift. The exhaustion of oil reserves, the change in biological and climatic factors and global warming are all themes associated with huge risks (since they entail the end of our societies) but in which the scientific knowledge that does exist is dominated by contradictory theories and uncertainties.

The new way in which industrial medicine is looking at the effects on (future) unborn generations of chronic exposure of this generation to harmful substances is another illustration of this development. The genetic effects on future generations are particularly feared in cases of chronic exposure to chemical substances. This is another area where science has little to offer. Putting the precautionary principle into practice is not very easy; the endless delays affecting the European REACH initiative[4] in the area of systematic studies of the toxicity of all new molecules clearly demonstrates this difficulty.

More globally, a society that has already been made largely safe fixes its gaze and makes its demands on increasingly distant time frames and focuses its efforts—in an echo of the foundations of individual biology—on surviving for as long as possible rather than simply on overcoming present limitations. It is also known that this "flight" of the safety horizon is particularly fragile and is rapidly adjusted in the presence of short-term threats.

To summary this fifth phase:

- The debate is shifting towards **allowing society to survive** in the long term and ensuring its future safety. The resilience of society is becoming more important than the safety of the individual.
- This debate sets out from the great, instrumentalised fears, from more or less general questions about the conditions under which society can survive and ends up increasing the immediate constraints on the function, production and structure of industries.
- The need is external to the industry, and it is virtual, huge, infinite and above all poorly defined, but due to societal pressures it cannot be avoided.
- People are more afraid of uncertainty than of major known risks. The future determines the present.
- There are no concrete ways to respond and culpability is certain, since even more than in the previous phase the efforts that are made will never be enough to overcome the danger.
- How does this final development impact on HOFs? In the immediate term these major fears are not expressed in the form of new methods, but the existence of these fears influences the way in which things are analysed and the resulting response. In this new development, it is not so much the trade-off between the various dimensions of risk that is important (as was the case in the previous phase), but the trade-off between the dimensions of uncertainty and change that have to be borne in mind when looking at feedback systems. The process of introducing and monitoring systems to pick up low noise signals (which are often rare and are seldom heard, or are even eliminated by current REX systems), the attention given by the press to whistle-blowers and their potential influence on future choices in relation to the organisation of work, are all firmly included within the scope of modern HOFs.

[4] REACH (Registration, Evaluation, Authorisation and restriction of Chemicals) is a European Regulation which came into force on 1 June 2007 and was intended to create a framework for registration, evaluation and certification of chemical substances and the restrictions applicable to those substances.

Conclusion

The five phases described above are like five blocks that have been left behind on a road that is being built.

It is not a total theory of safety that is being built here, still less a total theory of HOFs; it is the ongoing structure of society, its collective perceptions about its level of safety and the technological innovation that lies at the centre of the process, rendering some practices obsolete and automatically making it necessary to add and expand HOFs to deal with the new situations and products that appear on the market.

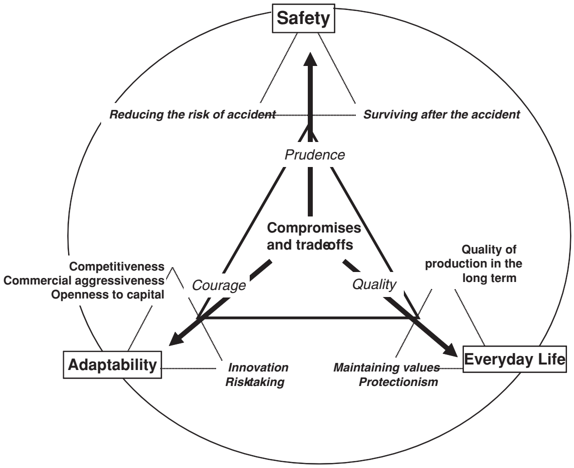
It is up to us to capitalise on the knowledge that has been gained about all these phases, without forgetting any of them, since there are still whole areas of industry that still rely on each one these phases. We have to lift our gaze high enough to see the global nature of these phenomena, address the problems in a sufficiently systematic way and maintain a sense of balance to counterbalance the numerous parochial disputes over methodological details or the wars of words over concepts which often have little long-term importance.

References

1. Taylor F (1911) The principles of scientific management. Harper & Row, New York. Reprint Norton Library, 1967
2. Pascarelli E, Quilter D (1994) Repetitive strain injury. Wiley
3. Mayo E (2003) The human problems of an industrial civilization. Contributors: first published 1933, reprint in Thompson K. The early sociology of management and organizations—Routlege, London. Publication Year: 2003
4. Gillespie R (1991) Manufacturing knowledge: a history of the Hawthorne experiments. Cambridge: Cambridge University Press, and Sonnenfeld JA (1985) Shedding light on the Hawthorne studies. J Occup Behav 6
5. Lewis E (1987) Introduction to reliability engineering. Wiley
6. Frankel E (1988) System reliability and risk analysis. Kluwer
7. Bell J, Holroyd J (2009) Review of human reliability assessment methods, report HSE, http://www.hse.gov.uk/research/rrpdf/rr679.pdf
8. Boy GA (2011) Handbook of human-machine interaction. A human-centered design approach. Ashgate, UK
9. Sarter N, Amalberti R (eds) (2000) Cognitive engineering in the aviation domain. Lawrence Erlbaum Associates, Hillsdale-New Jersey
10. Kanki B, Helmreich R, Anca J (2010) Crew resource management, 2nd edn. Academic Press, UK
11. Gollier C, Treich N (2003) Decision making under uncertainty: the economics of the precautionary principle. J Risk Uncertainty 27(1):77–103
12. Guilhou X, Lagadec P (2002) La fin du risque zéro, Ed Organisations

Chapter 5
Conclusion: The Golden Rules in Relation to Systemic Safety

Abstract In the complex interplay between opposing forces and their interactions, detailed knowledge of all the challenges facing the system (both safety and commercial challenges) and knowledge of its history, are vital in order to chart a safe course and make reasonable compromises and trade-offs. It is not really possible to make good decisions locally other than based on the criterion of overall trade-offs between these opposing forces; no decision in the area of risk management has any chance of being effective if there is no systemic vision.



The Enterprise is a System Incorporating Contradictory Tensions and Requiring Trade-Offs in the Area of Safety

When managing an enterprise, ongoing tensions must be managed in order to survive economically. The enterprise must guard itself against four risks:

- **not winning the contract**
  - Having no saleable products, competition, innovation
  - Difficulty making sales, inadequate distribution, economic recessions

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology,
DOI: 10.1007/978-94-007-6549-8_5, © The Author(s) 2013

- **inability to produce in time, to the expected quality and at the expected cost**
  - Quality of the production chain, image of the enterprise
  - Quality of maintenance
  - Good industrial relations
- **failure to control the financial support available for the business plan in the areas of innovation, production and sales**
  - The business model and the choice of company form
  - Cash, liquidity, borrowing, debt and investment
  - Partnerships, alliances and dependencies
- **inability to control the safety of production or of the product being sold**
  - Human disasters, the image of the enterprise
  - Exposure to penalties imposed by regulators

The management of these risks is distributed between different divisions: commercial, research, production and safety.

Each of these divisions tries to optimise its own road map, often to the detriment of the other divisions (challenge of resource sharing) and justifies its decisions on the basis of risks that constitute a threat to short-term or medium-term survival.

In this process of internal interaction, we know that trade-offs will spontaneously take place according to three constants:

- **the time that will elapse before the benefit or risk appears**
  - Priority is given to immediate benefits over hypothetical long-term benefits
  - Long-term risks are accepted in return for control over short-term risks
- **frequency versus severity**
  - Control over hypothetical severity is sacrificed in return for control over frequent, proven problems
- **salience and emergence versus rationality**
  - Internal trade-offs are prioritised in response to threatening external judgments (press, regulators, market, finances)
- **the time that will elapse before the benefit or risk appears**
  - Priority is given to immediate benefits over hypothetical long-term benefits
  - Long-term risks are accepted in return for control over short-term risks

On its own, this process of making trade-offs does not tend to promote safe solutions: these would reduce potential risks that are not immediate, with quite high costs (sometimes due to their own costs, particularly in personnel terms, time overheads for training, and would often slow down production due to more procedures and administration); the best advocates for these solutions are no doubt accidents that have already occurred… and the external regulators (if these exist) demanding compliance with regulations. Even this regulatory process, however, represents a potential trap, creating internal trade-offs in favour of safety solutions, since it often encourages the senior management to expect nothing beyond this compliance and monitoring of the indicators that are disclosed to regulators (whose priority is primarily to protect against industrial accidents) and the results will be less favourable unless adequate resources are made available for more systematic, in-depth actions. Finally, the message from the safety division also becomes confused with the quality process on the production line (whose management is often merged with the safety management): this function (which presents immediate challenges in terms of the image of the enterprise) is often given a higher priority than safety itself (which presents different challenges). This creates the illusion that the management is listening and prioritising proposals from the quality and safety division, even though they are only addressing points that have little to do with safety.

Ultimately the process of determining the internal priorities of the enterprise has to be negotiated within a space which allows only limited room for compromise and is subject to the following boundaries:

- the desire to reduce exposure to risks on one side;
- the desire to accept exposure to risk, in order to achieve secondary benefits that are considered more important, on the other.

There are three aspects to the art of intervening effectively in safety:

– gain and effectively communicate a systemic view of the risks facing the enterprise;
– do not resort to compromises on safety beyond a specific threshold;
– maintain the ability to manage the sacrifices that are made intentionally and are not part of the priority plan.

The Dimensions of Compromise and Offsetting Risks Within the Safety Division

We have just reviewed the key dimensions involved in offsetting risks within the enterprise. We will now see how to assess the margin for compromise within the intervention case/safety plan in order to prepare effectively for possible trade-offs during discussions at the strategic level.

The Three Essential Dimensions of Compromise and Offsetting in the Area of Safety

- The life cycle of the system. All industrial and service systems are born and die. The purpose of all interventions in safety is to extend the life of the system while providing the best conditions for ageing (healthy life, economic and physical health). This ambition is expressed in different ways at different phases of the system. The pressure on safety increases at the end of the cycle when the system has gradually exhausted its available margins for economic progress. As a result, the rules governing trade-offs in relation to safety actions (against economic constraints) vary at different phases in the cycle. Market factors and risk-taking are a higher priority than safety data during a large part of the cycle (when the safety model has to accept the risks that are taken and limit their negative consequences) but the paradigm goes into reverse at the end of the cycle, when the exposure to risk has been reduced through the actions taken and again takes priority over the economic model.

- Guidance for operators. If one wishes to ensure that the operator will continue to play an effective role in managing safety, four safety-related traits in the organisation of work must be prioritised. Each of these traits has its own control systems that must be adjusted according to the level of safety in the system: (1) regardless of the level of safety, do not count errors but instead count missed recoveries, (2) in the least safe systems, do not target prevention and simple compliance with procedures, but prioritise recovery strategies and damage limitation in training efforts, accident analysis and when designing procedures; it must be accepted that incidents will continue to occur and that it will not be possible to prevent them in future) and to progressively give the system more internal resilience to allow it to deal with such incidents, (3) in the safest systems, design a system of standards and constraints that is compatible with the desired performance of the enterprise. All excessive demands (which are no doubt satisfactory from the point of view of regulators and in terms of administrative conformity) will result in immediate violations. The enterprise will tolerate these and they will cause the system to start along a gradual and silent path towards loss of control, (4) for all systems, designing workplaces that can be understood intuitively by operators, without the need for excessive mobilisation of their cognitive resources, will allow most regular work to be done routinely, thereby freeing the operator's attention to focus on aspects that determine safety: anticipation, strategic orientation, choices and decisions.
- The economic model of the system. Not all systems face the same safety challenges. Public systems clearly have different demands from skilled trades, and unstable, highly innovative systems have different constraints. These different systems of work are based on different safety systems, each of which has its own way of internally managing the various compromises and trade-offs between safety and other dimensions within the enterprise. It is necessary to be able to recognise the type of system involved and to apply the rules to ad-hoc trade-offs. In a skilled trade type system, voluntarily seeking out exposure to new risks will be accepted and the safety card can be played by taking action to enhance the competence of the actors involved; in a traditional industrial system which is based on the HRO model, priority will be given to safety actions that concern the group and to procedures, leaving the system free to expose itself to risks and work under quite unstable conditions. In a public or ultra-safe system,

the safety model begins to take priority over the economic model and the attention and intensive work will be devoted to supervision and excluding possible exposure to risk.

One example of trade-offs that take place in medicine: the analysis of value [1, 2]. Why should we think about improving value rather than following a standard quality improvement process? The answer is simple: except in a few rare cases, improving quality has not delivered on its promises in the field of health. The results are poorly understood, disappointing and all too often in conflict with other competing realities, budgets or other priorities in delivering care. In fact a hospital must constantly resolve a range of contradictory forces: improving the quality of care, increasing the volume of care and reducing costs. Some hospitals manage to keep this impossible equation under control more successfully than others. The solution clearly seems to be an organisational one. Improving value is a response to this need. It means actively searching for the best compromise between these three systems which are in tension. Quality can and should be improved, but in real situations this can only be done by agreeing not to harm the other dimensions, or better still by making improvements in these other dimensions too. It should be noted that this approach does not contradict the fact that some improvements in quality, which do not result in savings, may be important and enjoy protection; these are relatively rare, however, in comparison with the quality solutions that do require trade-offs.

Patient safety in connection with errors and multiple organisational failures that shift professionals away from best practice (with multiple errors in management and inter-professional coordination) appears to be a typical area in which improvement in value has a major impact on the volume of care (poorly delivered care extends hospital stays and overloads the health care system), the cost of care and ultimately of course on quality and safety.

The process of improving value involves a compromise between the expected significance for the patient, the burden of implementation, the impact in terms of costs (which is hoped to be positive but is often negative) and the return on investment model.

Two examples of methods that can be used to calculate this compromise:

1. The profitability analysis model "cost—expenditure—saving" takes into account:
   - the cost resulting from the quality problem, specifically the estimated annual cost to the organisation of a quality deficit;
   - the expenditure that is approved to reduce the scale of the problem, i.e. the estimated resources that the organisation will have to invest in order to reduce the problem by 50 %, including expenditure associated with assessment of the problem and its evolution;

- clear savings/losses over 1 year, and in subsequent years, i.e. the annual cost resulting from the problem, from which the estimated expenditure is deducted for 1 year of improvement activity, to show how much the organisation could save in 1 year and in subsequent years.

Example: intervention intended to reduce pressures in a 600 bed hospital:

- cost of waste = $3 million, if reduced by 50 % = $1.5 million;
- expenditure approved for a 50 % reduction = $150,000: team time $30,000, training $70,000, mattresses $50,000;
- clear savings during the first year: $600,000 (supposing that it takes 6 months to design a plan and that this is then put in place during the following 6 months).
- Annual savings in subsequent years: $1.4 million.

2. The "cost of obtaining quality" method involves limiting the rolling out of quality to only those areas in which it has a real chance of resulting in gains. It is based on an estimate of the sum of the various real costs that may result from the quality process:

- prevention costs: costs resulting from activities intended to prevent quality failures;
- costs of assessing quality: costs of measuring and inspecting products or services to ensure conformity with quality standards (quality control);
- costs of internal failures (costs of failures affecting the product or service before delivery to the client);
- costs of external failures (costs resulting from failures after the service has been received by the client).

Five points are needed in order to build a value improvement process and make it succeed:

- a vision for the medical system that includes everyone, both patients and professionals:
- a targeted strategy that makes sense;
- a crescendo of action over 10–20 years to ensure that the organisation can make the changes needed to achieve the goal with a maximum of benefit: NOTHING CAN BE ACHIEVED IN THE VERY SHORT TERM during the first 2 years. IT IS NECESSARY TO BE ABLE TO INVEST AND PURSUE THE SAME STRATEGY while accepting that there will be a cost during the first few years (0–3 to 5 years), while the benefits will increase gradually;
- identification of the processes, their goals and the interactions between them;
- real, recognised, lasting leadership.

Two examples that have been studied particularly frequently due to their educational value:

1. The municipality of Jönköping in Sweden is an example known throughout the world for the quality of care delivered to its patients: the complete support which it enjoys among the public and professionals, the spectacular reduction in the number of SAIs, its effectiveness and the local care and support provided to all:

   – the common vision behind the process is simple and ambitious (the uniting slogan): "a good and attractive life for all in the municipality" (Note from the author: the vision does not even mention the medical aspect);
   – the strategy for improving quality is seen as a learning process rather than as a predefined method. A vigilant approach to new tools and new methods, which are systematically studied and tested and only adopted if they can be incorporated into the vision and the culture and will produce benefits;
   – duration is taken into account, as are stability and continuity within teams;
   – every aim that is addressed is prioritised and analysed in terms of added value, and is known and its value recognised by everyone from the secretary to the boss;
   – processes are identified and the support for their guidance, including financing, is assured. The General Manager is personally involved in piloting, as is the Innovation Manager and the Medical Director.

2. The Intermountain hospital system in the state of Utah in the United States (a non-profit organisation) offers another example of success in analysing value which has become known throughout the world:

   – the vision which is put forward and shared by everyone is that of "clinical excellence";
   – all the processes are prioritised and only the ones supported by the best clinical evidence (EBM) exists and offering the best economic value are retained. Eight specialities have been analysed and their guidelines have been drawn up along these lines, after a long process of learning about the method;
   – every clinical programme is the object of a full deployment project, incorporating skills in all sectors, medical of course but also in terms of administrative, IT and statistics. The results are monitored and reported continuously every month to make decisions on corrections or improvements (including deployment of new jobs or resources if necessary);
   – information sharing is at the heart of the commitment from professionals and patients (information technology, computerised medical records, a PC in every patient room, an intranet application to monitor clinical results, processes and outcomes);

- it took a few years to organise the system along these lines, but the current level of benefit is simply remarkable: the resources (both human and technical) are well above average for American hospitals, delivering highly superior value and service to patients and a healthy financial system resulting from the reinvestment of profits in innovation.

10 Golden Rules to Make an Intervention in Systematic Safety a Success

A review of the various points that are important in a systemic approach to risk within the enterprise reveals that there are 10 basic rules for effective management of the safety plan, its deployment, the sacrifices that have to be accepted and those that cannot be done without. The 10 rules can be divided up on three scales within the system: macro (the system), meso (the enterprise or hospital) and micro (the workplace).

At the MACRO Level

1. Do not run ahead of the demand for safety and seek to speed up the process of making a system safe: every system has a life cycle; its safety needs change at each stage; it is no use trying to offer a response that goes beyond the demand since that will only accelerate progress towards the end of the cycle.
2. Take into account the whole range of constraints faced by the enterprise. Accurately estimate the need for safety. The economic model and in particular the need for exposure to risk guides the safety model that will be chosen. Without seeking to change the economic model, it is necessary to prioritise the optimisation resources that exist within the chosen model before taking on board safety solutions from other models. Wider systemic analysis of the situation and the risk map may lead to the consideration of safety priorities or solutions that had previously been ignored because they were outside its scope (for example action in relation to obesity, prioritising a public health and food education policy rather than a priority focused on medical treatment).
3. Surviving accidents is just as important as being able to avoid them. The safety plan should not stop at preventing accidents. Just as error management at the individual level should leave room for detection and recovery, the management of safety in a complex system must leave room for management of an accident (crisis) so that the enterprise will survive the accident.

At the MESO Level

4. Design a “total” intervention at the macro, meso and micro level: No safety plan can limit itself to safety solutions that are delivered solely by front-line

actors. It is always necessary to envisage a part of the risk map and safety intervention process that specifically involves the executive, senior and middle management[12]

5. Pay particular attention to the roll-out of the plan among the middle management: feasibility, commitment and training. It must be remembered that risk management involves managing all kinds of problems that can destroy the enterprise. Safety, when it is understood in terms of avoiding accidents, is only one aspect of these risks, alongside economic risks, risks to the image, a lack of innovative know-how etc. In most cases it is actually the other risks that are most immediately perceived as threats to the enterprise. It is vital to be able to accept this as a relative priority while maintaining vigilance and an enlightened defence of pure safety aspects. It is particularly important to do well what can be done and even more important to have a good understanding of what one has (temporarily) decided not to do, in order to strengthen the protection in these areas. The management has a key role in relation to both functions. In fact the benefit of the safety plan, even if its extent is adjusted downwards during negotiations, can be obtained with the full support and commitment of the management. The safety plan should set out how the middle managers will be convinced, how the aspects that have been sacrificed will be explained to them and how those managers envisage the process of persuading and engaging front-line managers and front-line operators. The aspect of informing managers and operators

[1] For example: Carthey [3]. This system is based on an exploration using a questionnaire with 3 dimensions (the three Cs):

- commitment: a commitment by the managers to see the patient's safety as a priority, assessed on the basis of the decisions that are made, the trade-offs that actually take place in favour of safety and managers attending meetings about safety;
- commitment: a commitment by the managers to see the patient's safety as a priority, assessed on the basis of the decisions that are made, the trade-offs that actually take place in favour of safety and managers attending meetings about safety;—competence: the competence of the managers in the area of the safety of care and their training in this area;
- cognizance: awareness of risk: the actions (dashboards, crisis management) that are available to the management to be aware of the risks associated with care: non-punitive reporting, solutions to analyse risk collectively. How much time is reserved for noting and analysing these actions at divisional board meetings?

[2] Another article perfectly illustrates this systematic approach which is used by the health authorities in Scotland. The action goes by the name of NINEWELLS HOSPITAL and took place over 3 years. There are four main pillars: (1) establishing patient safety as a strategic priority, (2) establishing patient safety as a subject that is allocated the same amount of time and investment at management meetings as other subjects, (3), designing a sustainable organisation centered on patient safety in health care and in hospitals, with (4) a specific reflection at the institutional level in the area of education and training (professorial chairs, practical training courses etc.). The reforms that were carried out are monitored by measuring the rate of adverse events, using a method that is established for all health care institutions. The results specifically showed a reduction in mortality and a reduction in central venous line infections [4].

about what one has decided not to do in the safety plan is just as strategically important as the beliefs about doing well what one has decided to do.

6. Establish a fair analysis of the economic cost of incidents/accidents. The systematic analysis of safety and the trade-offs that it involves require a very realistic approach. Three sub-cases should accompany the risk analysis, in order to prepare for trade-offs and preserve as many as possible of the actions in this dimension of safety: one sub-case calculates the cost of quality losses and damages, a second offers an assessment of the impact on the commercial image of the loss of quality and mediocre safety performance, and a third calculates the potential impact of what one is probably not going to do, due to lack of resources or due to trade-offs in favour of other priorities.

At the MICRO Level

7. Establish a fair system for analysing the causes of accidents and incidents. A successful systematic approach must be honest and complete in its analysis of the causes of accidents and decisions to correct them. In particular this involves not excessively simplifying the causal link by only considering obvious errors by operators; equal priority must be given to actions to address latent errors in the organisation. Beyond these actions to address latent or patent errors, priority must also be given in the risk map for the risk of accidents associated with poor links between structures, without the need for any of the structures to have necessarily made mistakes itself (work on interfaces).
8. Allowing actors to maintain a fair degree of autonomy. The ultimate robustness or resilience of the system is always based on the remaining adaptive capacity of the actors. It is important to ensure that they are not strait-jacketed by unnecessary procedures, since each procedure slightly reduces the adaptive capacity of the system.
9. Putting in place a fair policy of incentives/checks and penalties. The desire to reduce errors and violations leads to a desire to maximise transparency while preserving a system of penalties for the most unacceptable cases. A consensus among the social partners on this system of transparency and sanctions is vital to the success of the safety plan. How will the system reward actors who submit reports, who will be involved (the Board and the management should be involved, and so should front-line actors), what are the real criteria defining what is unacceptable, particularly under suboptimal economic conditions in which the enterprise has to take more risks in order to survive? All these points must be included and discussed in the safety plan.
10. Creating a fair information system. Transparency should not only involve reporting adverse events. It should cover the whole process by which workers are informed about the strategies pursued by the enterprise, both from an economic perspective and in terms of safety policy.

The safety of a system is never a finished product; it is always in transition and it is perceived favourably or unfavourably by regulators, clients and by the enterprise itself.

There is no recipe for navigating safety (there is a toolbox available, but that does not deal with the problems). The level of control that can be achieved is based mainly on the ability to question oneself as a risk manager and understand the balances that must be maintained in order to survive today, and those that may evolve in order to give the enterprise a better chance of surviving tomorrow.

References

1. Ovretveit J, Staines A (2011) L'amélioration de la valeur dans les services de santé. Springer, Paris
2. Ovretveit J (2009) Saving through quality productivity improvement? a review of evidence of quality imporvments reducing costs to health service providers. The Health Foundation, London
3. Reason J, Carthey J, de Leval MR (2001) Diagnosing "vulnerable system syndrome": an essential prerequisite to effective risk management. Qual Health Care 10:ii21–ii25
4. Haraden C, Leitch J (2011) Scotland's successful national approach to improving patient safety in acute care. Health Aff 30(4):755–763

Index

R. Amalberti, *Navigating Safety*, SpringerBriefs in Applied Sciences and Technology,
DOI: 10.1007/978-94-007-6549-8, © The Author(s) 2013