



PERGAMON

Safety Science 37 (2001) 109–126

SAFETY SCIENCE

www.elsevier.com/locate/ssci

The paradoxes of almost totally safe transportation systems

R. Amalberti

Département des Sciences Cognitives, IMASSA, BP 73, 91223 Brétigny sur Orge, France

Abstract

Safety remains driven by a simple principle: complete elimination of technical breakdowns and human errors. This article tries to put this common sense approach back into perspective in the case of ultra-safe systems, where the safety record reaches the mythical barrier of one disastrous accident per 10 million events (10^{-7}). Three messages are delivered: (1) the solutions aimed at improving safety depend on the global safety level of the system. When safety improves, the solutions used to improve the safety record should not be further optimised; they must continue to be implemented at present level (to maintain the safety health obtained), and supplemented further by new solutions (addition rather than optimisation rationale); (2) the maintenance and linear optimisation of solutions having dwindling effectiveness can result in a series of paradoxes eventually replacing the system at risk and jeopardising the safety record obtained in the first place; and (3) after quickly reviewing ambiguities in the definition of human error and the development of research in this area, this article shows, through recent industrial examples and surveys, that errors play an essential role in the acquisition and effectiveness of safety, at individual as well as collective levels. A truly ecological theory of human error is developed. Theories of error highlight the negative effects of an over-extensive linear extrapolation of protection measures. Similarly, it is argued that accepting the limitation of technical systems performance through the presence of a minimum breakdown and incident ‘noise’ could enhance safety by limiting the risks accepted. New research opportunities are outlined at the end of this paper, notably in the framework of systems now safe or ultra-safe. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Ultra-safe systems; Human error reduction; Cognitive performance control; Optimised safety

E-mail address: rene-a@imaginet.fr (R. Amalberti).

1. Introduction: an implicit and general-purpose safety model

In most industries, safety remains governed by a few simple and self-explicit principles:

1. Conceptual designs generate systems with a high theoretical performance and safety potential, subject to technical and human failings. Breakdowns and human errors jeopardise operational and optimal safety, acting as ‘noise’ disturbing operations; ideally they should be totally eliminated or at least minimised. Conceptually, breakdowns and errors are symmetrically assessed, and detection logic, calculation of associated risks and safety objectives are supposed to be totally transferable from one design to the other.
2. Tools and machines are increasingly safe (fewer breakdowns); safety priorities thus increasingly aim at further reducing human errors (which still cause 60–70% of all accidents, if not more).
3. Reporting is fundamental to improve safety. Reporting describes the undesirable ‘noise’ plaguing the system in three areas of reference: the tool (machine) and its failures; the operator(s) and his failure(s); and the resulting situation (situational, organisational or systemic failings).

This common sense approach proved to be effective for decades, but is now beginning to lose relevance when optimising the safety of systems operating on the verge of total safety (Amalberti, 1997). This holds especially true for today’s ultra-safe macro-technical systems such as the nuclear industry, civil aviation or the European railroad system. The safety of these systems becomes asymptotic around a mythical frontier, placed somewhere around 5×10^{-7} risks of disastrous accident¹ per safety unit² in the system. As of today, no man–machine system has ever crossed this frontier; in fact, solutions now designed tend to have devious effects when systems border total safety. The ultimate safety solutions for automated aircraft have notably consisted in developing electronic cocoons and flight envelopes to protect the aircraft against pilot errors (excessive input on commands, under or overspeed, excessive bank and pitch angle). Quite paradoxically, these protections have become a significant cause of several recent incidents/accidents on glass cockpits, due to poor crew understanding of their software logic (incomprehensible mode reversion; Sarter and Woods, 1995; Abbott et al., 1996; Amalberti, 1998).

This article focuses on the example of these ultra-safe systems and first aims at demonstrating the limits of today’s safety approaches using only the linear extrapolation of known solutions. It then reviews developments in human error theories, the role played by these theories, and how they are handled in the safety policies of

¹ A ‘disastrous accident’ is an accident causing human death, and/or loss of property, and/or important consequences on the environment or on the system’s economic viability.

² Safety units vary according to industry or transportation mode (running time, passenger/km, number of airport movements, etc.). Figures in this article use statistics published by different industries, expressed in their specific unit.

systems which are now safe. The last part of this article explores possible solutions in the framework of these now ultra-safe systems.

2. An overwhelmingly perfect safety context, paradoxically subject to doubt

2.1. Safe and ultra-safe systems, and relative effectiveness of measures aimed at reducing breakdowns and human errors

Systems which are now ultra safe, i.e. where the risk of disastrous accident is below one accident per million events, behave differently safety wise than less safe systems.

They can be placed in the context of other systems:

1. Dangerous systems, where the risk of accident is greater than one accident per 1000 events, 10^{-3} (e.g. bungee jumping, or mountain climbing³). These are non-professional systems and fall outside the scope of this paper. They correspond to a personal quest for risk and thrills. Safety measures regulating these systems are highly individual.
2. Regulated systems, where the risk of accident lies between one accident per 1000 events and one per 100 000 events. Driving, chemical industries or chartered flights are examples lying in this category's upper safety bracket. Safety in these regulated systems is in the hands of professionals. Typical safety tools are: (1) regulations and procedures increasing hand in hand with safety performance; (2) accident or near-accident are almost repetition of stories of past accident and near-accidents; (3) error-resistant design (cutting down on the number of errors), and a reporting policy are dominant and efficient safety strategies; and (4) safety managers usually obtain results for newly implemented measures within a couple of years (in a 10^{-5} system); which means they normally get credit for their work.
3. Ultra-safe systems, where risk of disaster is below one accident per 100 000 or even one million safety units. Regularly scheduled civilian flights, railroads (in Europe) and the nuclear industry are examples of industries having reached this level. None of these systems has managed to achieve a global safety performance beyond one accident per 10 million safety units. These systems have specific features: (1) they tend to be ageing, are over-regulated⁴, rigid and

³ On serious expeditions in difficult, dangerous and unknown territory, 2–6% of participants die. This risk accumulates with repeated expeditions. Cumulative mortality in long-term expedition mountaineers ranges between 50 and 80%! (Source: O. OELZ, Risk assessment and risk management in high altitude climbing, 1999, 19th Myron. Laver International Post graduate course on Risk management, Department of Anesthesia, University of Basel, March 26–27, 1999, 5).

⁴ For example, the rate of production of new guidance materials and rules in the European Joint Aviation Regulations is significantly increasing while the global aviation safety remains for years on a plateau at 10^{-6} (over 200 new policies/guidance/rules per year). Since nobody knows really what rules/materials are really linked to the final safety level, the system is purely additive, and old rules and guidance material are never cleaned up. No surprise, regulations become inapplicable sometimes, and aviation field players exhibit more and more violations in reaction to this increasing legal pressure (Fig. 2).

highly unadaptive; (2) accidents are different in nature from those occurring in safe systems: in this case accidents usually occur in the absence of any serious breakdown or even of any serious error. They result from a combination of factors, none of which can alone cause an accident, or even a serious incident; therefore, these combinations remain difficult to detect and to recover using traditional safety analysis logic; (3) for the same reason, reporting becomes less relevant in predicting major disasters; and lastly (4) and not least, system managers work for their successors (over 8 years of inertia before being able to correctly assess the results of any new safety measure in the case of 5×10^{-7} systems⁵). The safety of these ultra-safe systems tends therefore to become a political rather than a scientific subject, with measures yielding visible results in the short term being favoured over long-term measures.

2.2. A few optimisation traps for ultra-safe systems

The insufficient definition of a number of concepts such as human error or incidents is not a problem as long as the system is not optimised. However, once optimised, this shortcoming can lead to misunderstandings.

2.2.1. Ambiguity in definitions

From the outset, the definition of human error seems implicit. However, this area is subject to ambiguity. In industry usually, only errors having non-acceptable consequences (i.e. outside the field of safe operations, as defined by procedures, instructions and safety analyses) are labelled ‘errors’. On the other hand, psychologists define error as an erroneous act, whatever its consequences, or the level at which it is detected and recovered.

The frequency rate between the two types of errors is probably around 1000 (one human error out of 1000 has unacceptable severe consequences in terms of safety⁶).

⁵ Consider the present annual departure rate equals to 20×10^6 , and suppose this rate remains stable. We have today an average accident rate of 1×10^{-6} , that turns out to be 20 catastrophies per year. The less the safety improves, the longer the time needed to be certain that this progress is not an artefact. With an error of the first kind of 5% (α risk to conclude that there is a change when there is not) and an error of the second kind of 80% (β risk which plans that, for 100 experiences (e.g. years), 80 at least will show the difference, if there is really a difference), you must wait for 2.28 years (from the time where the safety action is implemented) to conclude on a remarkable improvement of 50% of flight safety (a reduction of 10 accidents per year), but this time goes up to 32 years to conclude on a significant improvement of 15% (reduction of three accidents per year), which is an incredible target for civil aviation. If the improvement is only a reduction of one accident per year (19 instead of 20), the period to wait reaches 306 years.

The formula for this calculation is as follows:

$$n = (Z_{\alpha/2} - Z_{1-\beta})^2 / 2 \left(\text{Arcsin}\sqrt{P1} - \text{Arcsin}\sqrt{P2} \right)^2,$$

where $Z_{\alpha/2}$ is the risk of first order, $Z_{1-\beta}$ is the risk of second order, $P1$ is the present accident rate, and $P2$ the target rate, and n is the number of required observations of departures.

⁶ This figure comes from several inflight systematic observations by observers of crew errors (Amalberti and Wioland, 1997; Kemmler et al., 1998).

The amount of errors detected by the social and technical chain (workteam–procedures–protections) before the actual occurrence of a negative consequence is greater than 95%, and among remaining errors, 3–4% have no consequences on the system (Amalberti and Wioland, 1997). It is therefore obvious that, depending on the definition adopted, the implicit safety model (which presupposes the elimination of all errors) does not call for the same level of interpretation and requirements. The only solution to suppress all errors of front line actors is probably to suppress the front line actors (automation welcome), while the solution to suppress unacceptable consequences of human errors may take different directions, betting on situation control and recovery mechanisms instead of error avoidance.

The definition of incidents is less subject to ambiguity, but the granularity required for analysis and accident prediction remains a blurry concept. For a long time, incident analysis was only limited to serious and visible incidents requiring mandatory reporting. However, the notion of quasi-accident gradually evolved into the notion of quasi-incident, and databases mechanically stored an increasing number of reports on minor incidents. The Aviation Safety Reporting System database (ASRS), for example, which every year collects over 60,000 incidents deliberately reported by crews (nonpunitive policy towards the reporting of mistakes) perfectly illustrates this drift away from the original purpose (Website <http://olias.arc.nasa.gov/asrs/database.html>). This is a real drift, since all this additional information does not necessarily improve the prediction of future disasters. The logic behind the accumulation of data first relied on the strong predictability quasi-accidents had on accidents; extending the scope of safety analysis to quasi-accidents seemed natural. The same logic then applied by linear extrapolation to incidents, then to quasi-incidents, and eventually in turn to precursors of quasi-incidents. The result is a bloated and costly reporting system with not necessarily better predictability, but where everything can be found; this system is chronically diverted from its true calling (safety) to serve literary or technical causes. When a specific point needs to be proved, it is (always) possible to find confirming elements in these extra-large databases. For example, in the case of aviation, it has been possible in less than 10 years to supposedly prove that the leading causes of aeronautical incidents were communication problems, decision-making problems, time-pressure, organisational problems, automation or faulty situation awareness, each point being equally proven with highly convincing data (e.g. see numerous papers published in the proceedings of the Sixth, Seventh, Eighth, and Ninth Symposium on Aviation Psychology, Columbus, USA; most of these papers proceed by quoting convincing excerpts from incident/accident databases, showing local and supportive statistics).

The definition of operational safety is the last area subject to ambiguity, because it pursues two objectives that are not inevitably parallel. On the one hand, production safety is first and foremost designed to prevent the occurrence of events which may damage the production of goods or services. The production system must be protected against events which are usually minor, and which in all cases never fall in the realm of disasters (mistakes and breakdown of technical systems), but which regularly impair performance (technical failures, drifts in production, stoppage of machines). On the other hand, another area of safety focuses on accidents, with the

objective of protecting the system against disasters which may result from losing control of the situation, with severe damage to human beings (bodily injuries, deaths), equipment (severe degradation of production tooling), the company, the environment (ecological disaster with long-lasting consequences), society, and the economy (risk of bankruptcy). In this article, we are only interested by this latter area of safety, considering that the one and only unifying and intangible objective of safety must be the total absence of accidents. All other objectives (dealing with production safety) are secondary and aim at improving productivity rather than safety. These secondary objectives are legitimate for industry, even for ergonomics, but fall in another field of study with a different approach of problems. However, in certain cases optimising production safety objectives may jeopardise the optimisation of accident avoidance.⁷ This backdrop of safety strategies, ambiguities, and stagnation of results when dealing with ultra-safe systems echoes the limits encountered when analysing human reliability.

Human reliability analyses are almost entirely drawn from the work of Swain and Guttmann (1983) and their Technique for Human Error Rate Prediction (THERP) method. These analyses are still based on the description of human error, on understanding the consequences it has on systems, and on assessing its frequency rate. In more sophisticated analyses, the cause of error is also entered as an associated variable. This method proved to be sufficient and effective to manage safety in regulated systems, where risks remained fairly high. However it runs into difficulties when trying to assess the residual risk of ultra-safe systems, as the next section discusses.

3. Developments in research on human error

Research studies on human error are nothing new. However, many concepts were extensively reviewed during this century (see Fig. 1 for a summary). In psychophysics and behavioural psychology, scientists perceived error as an instrument helping measure the performance of experimental subjects. The study of error was not a goal per se, but just a mean to assess other psychological concepts. Threshold psychophysics were based on the delivery of the ‘good’ answer (correct perception of the stimulus/situation as defined by the experimentalist). In this domain the definition of error was never a problem, since the notion of a ‘good’ answer could not be

⁷ For example, when the region or the government have to show good results under the pressure of media during forest fires or flooding conditions, they ask for an oversignificant contribution of firemen, which increases local risks well beyond average, and turns out to jeopardize the firemen’s work. The rationale behind this conflict is that safety is not a unique matter. There are at least three permanent safety objectives in a large socio-technical system: one is the overall objective at the level of the company, or the top management. At that level, safety first aims at economically surviving in adverse conditions (whether it is caused by a production flaw, or a dramatic accident); a second objective is to manage production, and give priority to quality. The product is the first focus. The last objective is to protect individuals against physical and mental aggressions. Each level uses the other levels to achieve its own safety goal. Each level also resists the pressure coming from the other levels, in order to optimise and keep control of its own safety logic. The resulting overall macrosystem safety is an emergence of all these interactions.

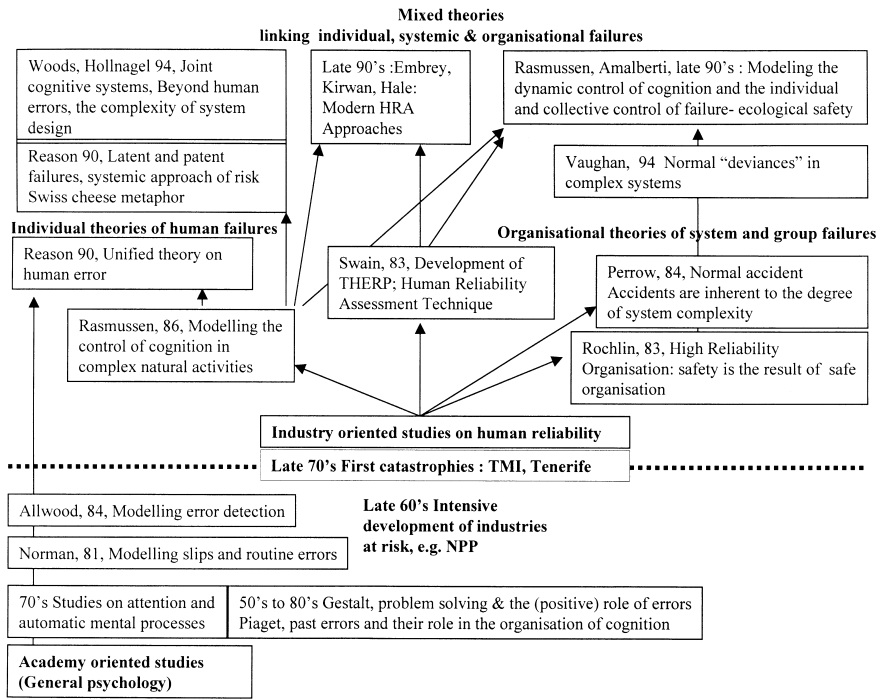


Fig. 1. Historical timeline showing main researchers and concepts related to human error.

discussed. Any deviation from the expected answer was automatically considered as an incorrect perception and an error (e.g. Green and Swets, 1966).

Since the beginning of the last century, another avenue has proved more fertile for research on human error. Gestaltists, the theoreticians of the good form mentioned very early on that errors are part and parcel of data reorganisation, which provides access to the solution (Einsicht, insight, Duncker, 1945). These same Gestaltists also reported that the system of generic solutions memorised by subjects are flagged by the memory of errors and failures associated to it (try not to do this or that), rather than by the memory of the successful course. Very early on, this current in psychology considered human error as having a beneficial and organisational role in cognition. The same logic recently inspired Dörner’s remarkable research on action control (Dörner, 1980, 1990). Dörner carried out numerous experiments with microworlds⁸ simulating complex situations inspired from real life. Two leading conclusions arise from his work:

1. On the one hand, reasons for failure are stereotyped. Subjects running into difficulties tend to escape into tried and tested solutions, setting aside difficult

⁸ Microworlds are simplified computer simulations of human complex activities at work, e.g. fire-fighting, air traffic control, or medical care.

points. They systematically and erroneously tend to carry out linear extrapolations, and never sufficiently take into account the collateral effects of the measures undertaken. Last and not least, they tend never to learn from their failures, or to identify their weaknesses and strongpoints; in a nutshell, they fail at (sufficiently) enhancing their meta-knowledge. These remarks also apply to failures in the guidance of very large systems. For example, Dörner often draws a parallel with the failures of war leaders, especially in Nazi Germany. These results can be remarkably applied to the paradoxes described in Section 1: introduction: even though safety no longer improves, safety managers still think of risk control in terms of linear extrapolations, and still apply the same old solutions (hunting down errors and failures, adding procedures) based on well-known recipes which help them feel secure (tried and tested values), without taking into consideration the collateral effects of these overstretched measures.

2. On the other hand, Dörner often underscores that training can correct this faulty approach. However, according to the author, training must first and foremost be aimed at obtaining a better awareness of the various types of existing errors, rather than at teaching general strategies designed to solve and manage complex situations.

A great number of these conclusions were regularly enhanced by other currents in research (e.g. Piaget, 1974), but the study of error never managed to achieve high priority in psychology before the end of the 1970s. Eventually, with the occurrence of the first major modern industrial disasters caused by human error (ground collision between two large aircraft in Tenerife, 1977, 587 casualties; and the well-known nuclear accident of Three Miles Island, 1979), the subject came to the forefront of industrial research.

Just as technical safety is improved through the reduction of technical breakdowns, it seems common sense to use a symmetrical rationale to improve safety through the reduction of human error. This human error reduction concept was extensively explored for some 20 years. Research funds were first spent on studying human reliability in engineering sciences, the human component being considered as an additional element in the system, similar to other technical components (i.e. refer to Swain and Guttman's work on human reliability assessment, 1983).

Research funds were also quickly earmarked towards life sciences and psychology, greatly enhancing psychological and psycho-sociological knowledge on the occurrence, typology and mechanism of human errors. At the end of the 1990s, research on human error was summed up and published in various reference papers and books (Norman, 1981,1988; NATO, 1983, symposium on human error; Swain and Guttman, 1983; Reason, 1990; Senders and Moray, 1991).

Four leading ideas, relatively new in the field of psychology and dealing with the regulation of safety and human reliability eventually came out of this industry-oriented research.

1. Mistakes are cognitively useful and cannot be totally eliminated. Psychology has acknowledged for a long time that errors act as flags on the learning curve.

Errors are more frequent with beginners, and decrease with experience. Learning also leads to automation of behaviour, resulting in an in-depth change in the nature of errors (Norman and Shallice, 1986). Errors cannot be merely assessed in terms of production. Activity consists partly in detecting and recovering errors. This detection recovery is so effective that Allwood (1984) considers it to be the true manifestation of expertise. In performance assessments, detection and recovery of failures is more important than the actual production of failures.

2. Mistakes made by individuals must be repositioned in a more systemic or sociologic framework to correctly analyse how they contribute to safety. The central objective is not error for its own sake, but the overall safety of the system. In order to explain mistakes and the risk it places on the system, it is necessary to reconsider the entire contribution of individuals to systems operations. Two main views resulted by this change in focus. The first view considers patent errors of front line actors revealing latent failures generated by design and organisation (Reason, 1990; Maurino et al., 1996). For this theory, the analysis of 'failing' players should not be restricted to front line players (drivers, field operators), but extended to all players involved at one time or another in the existence of a system plagued by latent failures which only require one or more additional error to occur. At the end, the overall system safety results from a sum of all in-depth defences (Reason's Swiss cheese metaphor). The second view considers individual failures to become catastrophic only if the organisation is not reliable; but the concept of safe organisation is not viewed as a sum of in-depth defences, but as a result of dynamic properties of the organisation, (Rochlin, 1993, High reliability Organisations).
3. Mistakes made by individuals are the consequences of the systematic migration of socio-technical systems towards augmented complexity, performance, and individual advantages. The more serious failures of large technical systems do not result from a single error or a single technical breakdown. They occur in circumstances laden with minor breakdowns and errors, which are almost normal in a context of increased pressure on production and fierce competition. Normal operations, where the system works at its highest productivity level, and which are tolerated by institutions even though they imply working at quasi-incident levels (deviance becomes a standard of normal operations), reduce the operational margin for recovery of minor failures, and result in brutal transitions or brutal migrations towards accidents and towards loss of control (Perrow, 1984; Wagenaar, 1986; Sagan, 1993; Vaughan, 1996; Rasmussen, 1997). Accident models change with high technologies and industrial competition, because systems now operate with reduced margins. From a psychological point of view, these notions of margins and cognitive functions operating at the limits of human performance were introduced by Rasmussen as early as the 1980s (Rasmussen, 1986, 1990). They were picked up and furthered by many authors (Hollnagel, 1993; Amalberti, 1996) who addressed the many complex problems encountered with the dynamic modelling of

cognitive control styles and of conscious error regulation, while also raising the fundamental question of the relationship between error and accident: are some errors more accident-prone than others?; Does the human operator take all his errors into account at the same level?; Does he identically defend himself vis-à-vis all errors?

4. With time, the body of knowledge discussed above gave rise to the idea that errors are a product of cognitive activity, regulated in the broader context of cognitive performance control. Fundamentally, an operator does not regulate the risk of error, he regulates a high performance objective at the lowest possible execution cost. In the human mind, error is a (necessary) component of this optimised performance result. The subject of psychological research is no longer the study of the actual error-producing mechanism, but the study of the cognitive mechanism used to control risk. Cognitive risk control is based on the dynamic control of a number of components including: being consciously aware of the scope of possible performance, of the difficulty entailed by the performance level required; choice of the style used to control activity (automatic/conscious); choice made to protect against errors, mechanisms involved to detect and recover errors; and, finally, tolerance to the production of a residual error rate. All these components can be comprehensively regulated through learning (gradual development of a repository of risk control styles) and reconsidered in real time during the activity. The resulting balance is a so-called natural or ecological safety, which does not aim at suppressing all errors, but at controlling them within an acceptable margin (Amalberti, 1996).

Recent studies (Amalberti and Wioland, 1997; Wioland, 1997; Plat and Amalberti, 2000) show how error control is just another variable among many others within the superseding control called situation control, which itself rests under the comprehensive control of dynamic cognition, often called cognitive control mode (Hollnagel, 1998). The results of these experiments on microworlds and on real-life situations can be summed up in a number of significant findings:

1. Subjects produce a more or less constant rate of error, whatever their expertise, except for absolute beginners. The average rate of error observed in various situations hardly ever varies, and stands at one to three errors per hour, according to the situation and the challenges at hand. The number of errors tends to decrease in more demanding situations (increased cognitive control), but at the same time the recovery rate also tends to collapse (lack of sufficient resources for on-line control and recovery).
2. The nature of errors changes with expertise. Routine-based errors increase with expertise, whereas knowledge-based errors decrease.
3. This error flow stays under control. Seventy-five to 85% of errors are detected, with a higher detection rate for routine-based errors than knowledge-based ones. Notably, expert subjects tend to disregard an increasing number of errors having no consequences on the work underway (28% of non-recovered errors are considered as inconsequential by expert subjects, whereas this

percentage is only 8% in the case of beginners).⁹ This last result strongly indicates that the criterion used to regulate the system is whether they feel that the situation is under control rather than merely counting the number of errors.

4. Meta-knowledge and confidence (which, by the way, are closely related notions Valot and Amalberti, 1992) lie at the core of cognitive risk control. Using these results and with reference to the concept of “field of safe operations” developed by Gibson and Crook (in Flach et al., 1994) and furthered by Rasmussen (1997), the subjects’ performance bracket can be modelled, with higher and lower limits. This performance bracket is characterised by mute areas (fields of safe operations, where the subject believes he is in control of the situation) and turbulent areas, bordering the zone where the situation gets out of hand and where the subject receives cognitive alarm signals telling him he is increasingly running the risk of losing control. These signals are a logic target for various safety measures which, rather than trying to eliminate errors, tend to increase the spontaneous cognitive control within the performance bracket (by strengthening signals). However, these cognitive signals should never be masked (which unfortunately is the case with a number of automated devices inducing over confidence), and should provide the operator with ways to really control his performance bracket (allowing him to downgrade or upgrade his performance level when limits are approached) (Rasmussen and Vicente, 1989).

In summary, research on human error first tried to understand why errors occur. The rationale behind mistakes greatly helped improve the safety of systems whose frequency of disastrous errors (i.e. leading to system loss) is above 10^{-5} . Actions were then undertaken to prevent errors, using passive ‘error-proof’ systems and active solutions, where training and education played a major role.

These solutions always prove to be helpful in the case of systems under the safety threshold of 10^{-5} . Beyond that limit and driven by technology and the never-ending quest for safety, research on error focused on the consequences of a total or almost total elimination of human error, either by replacing operators with an automated system or by controlling their performance through the strict application of procedures.

During the course of this research, error was found to play an ecological role in the control of performance; another finding was that the flow of error is regulated by cognition, and that detected errors play a central role in the maintenance of situation awareness. Automated solutions weaken situation awareness, which is probably why their effectiveness starts wearing out around 10^{-6} .

Beyond 10^{-6} , safety optimisation increasingly depends on strengthening the ecological mechanisms of cognitive error regulation rather than on fighting them. The

⁹ This result has been duplicated in a large cockpit safety survey made by Lufthansa where it has been shown that 43% of inflight human failures sorted themselves out, or were intentionally late recovered by crews (Kemmler et al., 1998. Analysis of inflight situations and development of preventives measures. Paper presented at the CRM’s managers conference, Frankfurt, 2 November).

principles regulating the co-operation between external aids and the mechanism spontaneously developed by operators are addressed in the last paragraph.

4. Extending the ecological safety model to incidents

Breakdowns and incidents follow the same logic as errors. An incident-free system becomes mute, and its safety can no longer be tuned. Investments stop being directed at safety and are earmarked towards improving performance; the control and the spontaneous limitation induced by the occurrence of incidents no longer play their role. The system can then brutally find itself in a situation of disastrous accident because its over-stretched performance has given rise to new risks (Rasmussen, 1993).

Risks arising from increasing performance levels beyond a certain threshold come on the one hand from the reduction of margins and recovery opportunities in degraded conditions, and on the other hand from the more serious nature of accidents (more consequences on environment, people and property). In other words, the presence of incidents reduces ambitions in system performance (Argyris, 1990). Beyond a certain incident-reduction quota, the absence of incidents (known or visible), as opposed to the presence of a minimum number of incidents, does not prevent disastrous accidents from occurring; it may actually even increase the risk of accident (the system gets out of control because of over performance). Safety and performance tend only to be harmoniously balanced within a limited bracket.

Optimum safety is obtained through the careful monitoring (and the tolerance) of a minimum number of incidents (the number reached when the system is balanced around the mythical barrier of 5×10^{-7}). Specific monitoring techniques used to track this minimum level of incidents help ensure an optimal safety level and are described in the last paragraph.

5. Solutions and perspectives

5.1. General development model for large systems

This article started with an obvious statement: errors and breakdowns must be eliminated to increase safety. This paradigm holds true to optimise the safety of man-machine systems up to one risk of disastrous failure for 100,000 events, (1×10^{-5}). To reach this level of effectiveness, definitions do not need to be overly clarified, nor are complex cognitive processes required to come into play. Increased procedures, improved staff training, automation, and conventional error blocking solutions are effective. Traditional safety analysis is an adequate tool at this safety level, even though allocation of failures to categories may remain quite approximative.

To go beyond 10^{-5} , the safety approach must necessarily clarify ambiguous definitions and call on more sophisticated cognitive concepts. Solutions previously used

can be further optimised up to 10^{-6} , but they tend to block the system’s adaptive capabilities, and to have devious effects which may actually make it difficult to maintain safety at the edge (Fig. 2).

Optimising safety around 10^{-7} requires priority to be given to safety rather than to performance improvement, and a different attitude to be adopted vis-à-vis mistakes and breakdowns. The solution is (1) to maintain solutions which have helped reach previous safety levels, without over optimising them (reduction of human errors and incidents, specific treatment of systems logic), and (2) to additionally implement new safety strategies (described below).

It may be impossible today to go beyond this limit. Today’s systems may not be able to reach safety levels higher than this record of 10^{-7} . Most of today’s

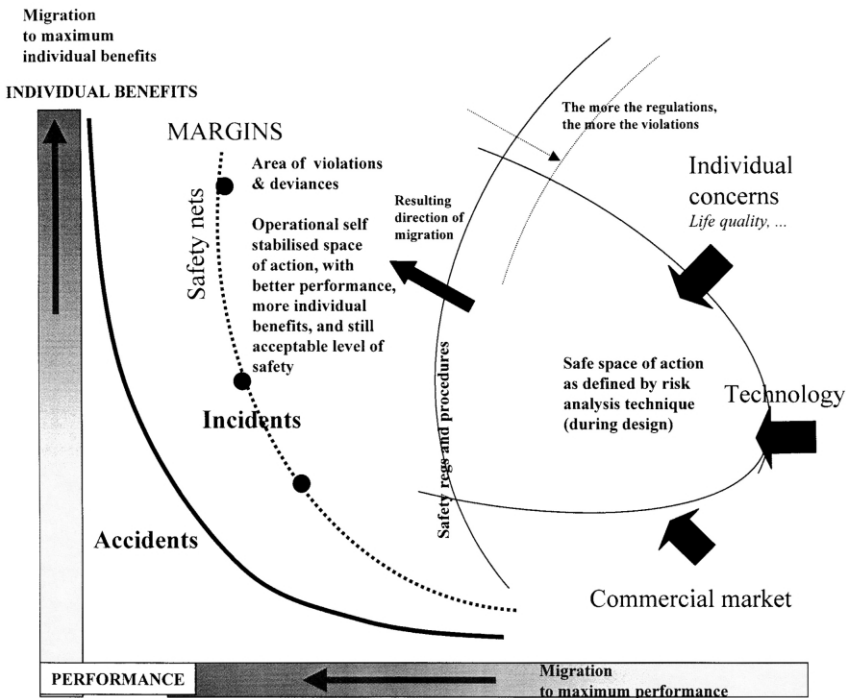


Fig. 2. The trap of over regulation. The safe space of performance, as expected and calculated during design, is contained within three boundaries: the individual and social regulations; the market rules; and the safety rules. When in use, the system migrates through the safety boundaries towards more performance and more individual benefits (see Rasmussen, 1997, for development of this idea). The new resulting operational space of performance becomes largely positioned outside the initial safe space of performance. This new space is characterised by reduced margins to incidents and accidents (despite safety remains acceptable) and numerous violations and deviance. The safety trap should consist in that situation to continue a simple-minded strategy fighting violations with the development of new regulations. Cumulative regulations will then have the effect to change nothing in operations (the system is stabilised), mechanically increase violations, increase reluctance and opacity in incident reporting, and add noise in the safety monitoring strategy.

man–machine systems were designed in the 1960s. These systems are over-optimised today, but their operating logic remains identical to their original design with a man or a system of men placed at the heart of their normal or degraded regulation mechanism. This design contains its own limitations. No system will last forever, and we are probably dealing today with ageing logic which will someday be replaced by a different logic (i.e. complete datalink for aviation, or automated nuclear plants) once the technology is finally mature, and probably after the occurrence of a highly publicised major disaster. This new logic will in turn be optimised over time (20, 30 or 40 years), and may also eventually reach its maximum safety potential, probably beyond today's 10^{-7} .

To sum things up:

1. the main operating principles of systems bear within themselves a maximum safety potential;
2. to go beyond this potential, it is necessary to change the nature of the system (eternal technological cycle);
3. safety measures are different according to the system's optimisation level; and
4. at the end of optimisation (ageing system), when safety has nearly reached its maximum level, the objective is no longer improving but containing the system within its optimised limits as long as possible; a comparison with geriatrics can perfectly illustrate this point: very old persons are not medically cared for in order to be cured using intensive treatment, but only to prevent suffering and to prolong life with maximum effectiveness; the drugs used are less potent and interactions between different ailing organs need to be taken more seriously than with younger patients (avoid treating the liver by killing the heart).

5.2. *How to ensure that the optimized safety of ultra-safe systems remains on its toes*

Three new safety priorities need to be set to maintain safety effectiveness in ultra-safe systems:

1. The systemic objective is not to entirely eliminate all human errors, incidents or even accidents. The system must be allowed to age ultra-safely, and all efforts must be made to protect it against the accident which would prove its doom ('the big one'). This accident must be avoided at all cost and protection policies can lead to specific safety measures, such as local safety reinforcements in high-risk situations. Hybrid systems making use of technological leaps may even be developed with improvements in technology earmarked towards the areas where the reactivity of media and people could make it impossible to have any (media and legal) control over the consequences of an accident.
2. To maintain system safety on the asymptotic edge of 10^{-7} , defence strategies need to be aggregated, but no single strategy should be overly and unreasonably optimised. Over optimisation freezes adaptive capabilities of human and technical systems, while covering up minor system failures. A system can be ultra safe only if its failures are minor, known, circulated and used as booster

shots and performance restrictors. Since minor failures do not usually predict any kind of accident, the system must be monitored through the global monitoring of its volume of failures. When the volume increases, accident is looming by, even if accident modalities remain almost totally unpredictable. The link between failures and accident being more mass-related than analytical, the observation of a significant number of failures should not (automatically or a priori) require each failure to be addressed by implementing selective barriers. More generally, the increase in the number of failures results from the upstream emergence of new systemic causes, which need to be found and addressed: new risk generators, poorly controlled deregulation processes, economic crises, etc. (see Reason, 1990; or Woods et al., 1994, for the development of this analysis concept called “going beyond error”). Once again, this logic specifically applies to ultra-safe systems. Incident volume monitoring is probably less relevant for less safe systems (actually, it has relevance but a direct analysis of incidents proves even more relevant in this case). Maintaining safety at the edge also requires, in parallel, the maintenance of more traditional safety strategies, such as error-resistant systems design, or procedure-based operations. The message is, however, that over optimising these latter solutions cannot help improve safety beyond 10^{-6} , because over-optimised measures risk being counterproductive, by over rigidifying the system (elimination of adaptive flexibility; Rasmussen, 1997). Better safety results could perhaps be obtained by supplementing these traditional solutions with new solutions aiming at improving the global ecological cognitive control of the situation, eventually helping reach 10^{-7} , without imposing a systematic and immediate error recovery, and strict adherence to procedures.

3. The way errors are taken into account in the design process must also change. Routine-based errors, knowledge-based errors and violations must be handled differently:
 - Routine errors are frequent and almost infinite in number, but tend to cause finally less accident than faults, thanks to the efficient associated mental recovery mechanisms (Allwood, 1984; Van Der Schaaf, 1999). Analysing their root causes has therefore little relevance to improve safety, unlike the analysis of their propagation. The problem of ergonomics and safety is that all routine-based errors should not be blocked off, since they flag the operator’s expertise (Riso et al., 1987; Visciola et al., 1992). However, safety requires the propagation of these errors to be blocked off through an adequate process, notably aimed at increasing their detectability, or at multiplying safety nets (but this can cause its own safety problem, as mentioned earlier in the paper; the net gain is positive, but usually not as much as anticipated).
 - Errors caused by misunderstandings must be addressed differently, since they usually highlight difficulties linked to the technical system architecture (Woods et al., 1994). Contrary to routine errors, these errors require root cause analysis, in order to find adequate remedies, which generally involve

reviewing in-depth core logic, or providing better instructions and better training to the staff.

- Violations are more complex. Safety services often (always?) forbid them strictly, but their analysis is not devoid of ambiguity. Systems and senior management often tolerate a number of violations because they allow the work to be done, given various adaptations (a machine tool can be manned by a worker having a lesser skill level than expected in the first place, a shift team can work with one less operator, a plane which broke down can be granted a take-off waiver to be flown back to a place where it can be more easily repaired, etc.). Other violations are tolerated because they contribute to improving industrial relations (Girin and Grosjean, 1996; Vaughan, 1996). Halo effects are commonplace, violations usually calling for new violations by extending tolerances and common practices. Paradoxically, violations are also more numerous in ultra-safe and ultra-regulated systems, because system rigidity requires more numerous adaptive processes to manage work constraints. Of course, this does not justify violations, but must be taken into account in the attitude-adopted vis-à-vis these errors. Safety approaches must be based on making these violations visible, and on controlling them, rather than aiming at their total elimination. Violation control requires, above all, the establishment of a corporate safety culture, with common values regulated and protected by a true dialogue among the various reporting levels and the implementation of successive safety nets.

6. Conclusion

Ultra-safe systems have reached today's safety level through a lengthy, extensive and crisis-ridden optimisation process, during which these systems have aged and matured. However, this process has also made systems more fragile and less adaptive. These systems will eventually be replaced by others having a different safety potential; thus goes the development of technical cycles. However, the challenge for today's safety practitioners is to manage the transition from these older technologies to innovative ones, under the best possible economic, human and technological conditions.

A conservative attitude, by which an older system with a high safety level is kept on going may be a possible solution (to avoid employment problems, to achieve a more economical transition with other techniques, or to ensure the optimisation of gains now reaped with the present system). The message in this article is threefold; (1) it is essential in such a case to fully understand that in this area of operation, system behaviour becomes special; (2) such a system must be treated with methods allowing it to remain simultaneously at the edge of safety and at a sufficient performance and competitiveness level to resist market constraints; and (3) but it is also important to recognise that these systems are nearing the end of their life, and should not be placed off-balance by requiring operations to take place within unreachable performance and safety objectives.

In conclusion, these systems provide remarkable opportunities for safety and human error research. They pose new questions, require the development of new research paradigms, encourage cross-fertilisation in this field of research, and will eventually contribute to the development of improved safety levels for the new systems to come.

7. Disclaimer

The ideas expressed in this paper only reflect the opinion of the author and must not be considered as official views from any national or international authorities or official bodies to which the author belongs.

References

- Abbott, K., Slotte, S., Stimson, D. (Eds.), (1996, June). *The Interfaces Between Flightcrews and Modern Flight Deck Systems* (Report of the FAA HF Team, June 1996). FAA, Washington, DC.
- Allwood, C.-M., 1984. Error detection processes in statistical problem solving. *Cognitive science* 8, 413–437.
- Amalberti, R., 1996. *La conduite des systèmes à risques*, PUF, Paris [The control of systems at risk].
- Amalberti, R., 1997. Paradoxes aux confins de la sécurité absolue. *Annales Des Mines* Fev97, 9–15. [Paradoxes of absolute safety within the limits of science.]
- Amalberti, R., 1998. Automation in aviation: a human factors perspective. In: Garland, D., Wise, J., Hopkin, D. (Eds.), *Aviation Human Factors*. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 173–192.
- Amalberti, R., Wioland, L., 1997. Human error in aviation. invited paper to the International Aviation Safety Conference 1997 (Iasc-97). Rotterdam Airport, The Netherlands. In: Soekkha, H., (Ed.), *Aviation Safety*, Utrecht: Vsp, pp. 91–108.
- Argyris, C., 1990. *Overcoming Organisational Defenses*, Prentice Hall, Englewood Cliffs, NJ.
- Dörner, D., 1980. On the difficulties people have in dealing with difficulty. *Simulation & Games* 11 (1), 87–106.
- Dörner, D., 1990. The logic of failure, *Phil.Trans. R. Soc. London*, B327, 462–473.
- Duncker, K., 1945. On problem-solving. *Psychol. Monographs*, 58 (whole no. 270).
- Flach, J., Hancock, P., Caird, J., Vicente, K. (Eds.), 1994. *Ecology of Human Machine Systems: A Global Perspective*. Lawrence Erlbaum Associates, Hillsdale NJ.
- Girin, J., Grosjean, M., 1996. *La transgression des règles au travail*. L'harmattan, Paris. [Rules Transgression at Work.]
- Green, D., Swets, J., 1966. *Signal Detection Theory and Psychophysics*, Wiley, New York.
- Hollnagel, E., 1993. *Human Reliability Analysis, Context and Control*, Academic Press, London.
- Hollnagel, E., 1998. *Cognitive Reliability and Error Analysis Method, CREAM*, Elsevier, North Holland, London.
- Kemmler, R., Braun, P., Neb, H., 1998. Analysis of inflight situations and development of preventives measure. Paper presented at the CRM's Managers Conference, Frankfurt, 2nd November.
- Maurino, D., Reason, J., Johnston, N., Lee, R., 1995. *Beyond Aviation Human Factors*, Ashagate-Avebury Aviation, Aldershot, UK.
- NATO, 1993. *Advanced Research Workshop on Human Error*, Bellagio, Italy.
- Norman, D., 1981. Categorization of action slips. *Psychological review* 88, 1–15.
- Norman, D., 1988. *The Psychology of Everyday Things*, Basic Books, New York.

- Norman, D., Shallice, T., 1986. Attention to action: willed and automatic control of behaviour. In: Davidson, R., Schwartz, G., Shapiro, D. (Eds.), *Consciousness and Self Regulation: Advances in Research*. Plenum Press, New York, pp. 1–18.
- Perrow, C., 1984. *Normal Accidents, Living With High Risk Technologies*, Basic Books, New York.
- Piaget, J., 1974. *La prise de conscience*, PUF, Paris [The Emergence of Consciousness].
- Plat, M., Amalberti, R., 2000. Experimental crew training to surprises. In: Sarter, N., Amalberti, R. (Eds.), *Cognitive Engineering in the Aviation Domain*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Rasmussen, J., 1986. *Information Processing and Human-machine Interaction*, Elsevier North Holland, Amsterdam, pp. 165–187.
- Rasmussen, J., 1990. Human error in organizing behavior. *Ergonomics* 33 (10/11), 1185–1190.
- Rasmussen, J., 1993. Learning from experience? How? Some research issues in industrial risk management. In: Wilpert, B., Qvale, T. (Eds.), *Reliability and Safety in Hazardous Work Systems*. Springer Verlag, Berlin, pp. 43–66.
- Rasmussen, J., 1997. Risk management in a dynamic society, a modelling problem. *Safety science* 27 (2–3), 183–214.
- Reason, J., 1990. *Human error*, Cambridge University Press, Cambridge, UK.
- Riso, A., Bagnara, S., Visciola, M., 1987. Human error detection process. *International Journal Man-Machine Studies* 27, 555–570.
- Rochlin, G., 1993. Essential friction: error control in organisational behaviour. In: Akerman, N. (Ed.), *The Necessity of Friction*. Springer/Physica Verlag, Berlin, pp. 196–234.
- Sagan, S., 1993. *The Limits of Safety: Organizations, Accidents, and Nuclear Weapons*, Princeton University Press, Princeton, USA.
- Sarter, N.B., Woods, D.D., 1995. “How in the world did we ever get into that mode?” Mode error and awareness in supervisory control. *Human Factors* 37 (1), 5–19.
- Senders, J., Moray, N., 1991. *Human Error: Cause, Prediction and Reduction*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- Swain, D., Guttman, H.E., 1983. *Handbook Of Reliability Analysis With Emphasis On Nuclear Plant Applications*, Nuclear Regulatory Commission, Nureg/Cr-1278, Washington DC, USA.
- Valot, C., Amalberti, R., 1992. Metaknowledge for time and reliability. *Reliability Engineering and Systems Safety* 36, 199–206.
- Van Der Schaaf, T., 1999. Human recovery of errors in man-machine systems. *Proceedings CSAPC 99*, Villeneuve d’Asq: France, 21–26 September.
- Vaughan, D., 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*, University of Chicago Press, USA.
- Visciola, M., Armandi, A., Bagnara, S., 1992. Communication patterns and errors in flight simulation. *Reliability Engineering System Safety* 36, 253–259.
- Wagenaar, W., 1986. The causes of impossible accidents. The Sixth Duijker Lecture, University of Amsterdam.
- Wioland, L., 1997. *Etude des mécanismes de protection et de détection des erreurs, contribution à un modèle de sécurité écologique*, Thèse de doctorat de psychologie des processus cognitifs, Université Paris V, Décembre 1997. [Study of error protection and detection mechanisms: contribution to an ecological safety model.]
- Woods, D., Johannesen, D., Cook, R., Sarter, N., 1994. *Behind Human Error*, CSERIAC. Wright Patterson Air Force Base, OH.