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,

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

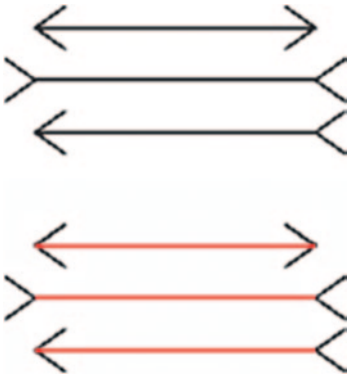


Fig. 2.1 The observer states that the line with the points facing towards the middle is longer

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.

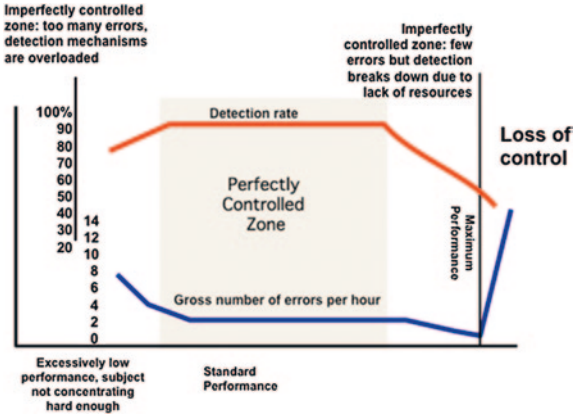
Under more attentive conditions, the average error rate is closer to two errors per hour (the rate observed in civil aviation over a series of more than 3,000 flights; these results were obtained from the Line Oriented Safety Audit (LOSA) type of large-scale online audit techniques [19, 20]).

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

http://www.springer.com/978-94-007-6548-1

Navigating Safety
Necessary Compromises and Trade-Offs - Theory and Practice
Amalberti, R.
2013, XV, 132 p. 7 illus., 6 illus. in color., Softcover
ISBN: 978-94-007-6548-1