Journal of Risk Research
Vol. 10, No. 7, 959–975, October 2007

Routledge
Taylor & Francis Group

ARTICLE

Errors and Failures: Towards a New Safety Paradigm

CLAUDE GILBERT*, RENE AMALBERTI**, HERVE LAROCHE[†] & JEAN PARIES[‡]

**CNRS, PACTE-Politiques publiques, Actions politiques, Territoires, Institut d'Etudes Politiques de Grenoble, France; **Hôpital Militaire du val de Grâce, IMASSA Institut de Médecine du Service de Santé des Armées, France; [†]ESCP-EAP, Ecole Supérieure de Commerce de Paris, European School of Management, France; [‡]Société Dédale, France*

ABSTRACT In France studies on technological risks began to question errors, failures and vulnerabilities at the end of the 1970s, focusing mostly on analyzing major accidents as consequences of the increasing complexity of socio-technical systems. During the 1980s and 1990s, research studies carried out in different fields (industrial risks, natural risks, health risks) underlined the importance of organizational factors in system vulnerabilities. Still, the bases of safety policies and safety management remained unchanged, with a strong reliance on rules and procedures. Building on an interdisciplinary reflection carried out at the beginning of the 2000s, this paper calls into question the prevailing approach as regards safety. Identifying the basic assumptions behind safety policies, it is argued that, in light of research advances in various fields of safety studies – and more specifically in cognitive ergonomics – they appear to be basically flawed. In a quite radical manner, a recognition of errors and failures as a part of the usual functioning of socio-technical systems, which are "naturally" unstable systems, is called for. As for risk control, it appears to result mainly from the capacity of operators, working groups and organisations for dynamically "making up" for errors and failures. These analyses open very stimulating prospects of research. However, the question of their social and political acceptability must be seriously considered.

KEY WORDS: Safety, error, failure, risk, sufficiency, safety paradigm

Correspondence Address: Claude Gilbert, CNRS, PACTE-Politiques publiques, Actions politiques, Territoires, Institut d'Etudes Politiques de Grenoble – CNRS. Tel.: 33 (0)4 76 82 73 02; Fax: 33 (0)4 76 82 73 04; Email: Claude.Gilbert@msh-alpes.prd.fr

1366-9877 Print/1466-4461 Online/07/070959–17 © 2007 Taylor & Francis
DOI: 10.1080/13669870701504764

Introduction

In the late 1970s and the 1980s in France, reflection in the human and social sciences on risks was related to the emergence and recognition of 'major risks' (Lagadec, 1981), associated with the growing complexity of certain dangerous activities (e.g., chemicals, nuclear energy). During this period researchers' attention focused primarily on crisis situations triggered by these activities (Lagadec, 1981, 1988). The analysis of the conditions in which accidents and disasters actually occur and, more generally, of the reliability and vulnerability of large socio-technical systems, was left to the 'hard sciences' for a long time (primarily the engineering sciences). To a large extent the human and social sciences were satisfied with general research on the 'reliability of organizations' (La Porte, 1994, 1996, 2001; Rochlin, 1991, 1996, 2001), or studies that linked failures to the complexity of systems, sometimes with brief references to the idea of 'normal accidents' (Perrow, 1994), or else that referred to non-compliance with security rules (Girin and Grosjean, 1996). There was little research that focused explicitly on the conditions in which accidents and disasters had occurred in firms and organizations managing potentially dangerous activities. It was therefore against the flow that reflection on errors and failures was launched.

This reflection was nourished above all in the late 1990s and early 2000s by the dissemination, by the RCSC Programme, of Anglo-American and North European studies formerly unknown beyond a limited circle of researchers. It was then developed in the framework of the series of seminars on the topic *Le risque de défaillance et son contrôle par les individus et les organisations dans les activités à hauts risques*[1] (2000–2003) organized jointly by the research programme *Risques Collectifs et Situations de Crise* (RCSC) managed by the CNRS and the *Action Concertée Incitative 'Cognitique'* at the ministry of research. These seminars covered various disciplines, including sociology, political science, management science, information science and technology, psychology and cognitive ergonomics, engineering science and neuroscience, among others.

The objectives were: to identify the implicit assumptions in dominant approaches and, over and above disciplinary divisions and the weight of demands placed on experts, to reach an overall understanding of problems of safety; to understand modern accidents by adopting a position that was increasingly one of comprehending the "normal" industrial risk management; to analyse errors by adopting a position at multiple levels (individual, working group, organizational, etc.) and by 'taking into account multiple aspects (technical, human, organizational) so that problems of a cognitive nature can be situated and contextualized.

Our first observation concerns the existence of a strongly dominant safety paradigm in which the management of hazardous activities (in industry – including nuclear – transport, major networks, hospitals, etc.) is

[1] The risk of failure and its control by individuals and organizations in high-risk activities.

imbued by a particular view of safety. This view is usually found among control authorities directly in charge of high-risk activities and is supposed to meet the expectations of the other actors and the populations concerned. It relies on a set of six basic assumptions:

1) People's safety (and, to a lesser degree, that of equipments and the environment) is an absolute priority.
2) All available means – in terms of knowledge and action – have to be implemented to guarantee this safety through risk prevention in dangerous activities.
3) The mobilization and instrumentation of science and techniques make it possible to identify and define most of the risks and thus to envisage a possible risk control.
4) Knowledge of risks serves as a foundation for the elaboration of norms, rules and procedures which can be used to frame dangerous activities and to manage them.
5) Safety in dangerous activities depends on the avoidance and elimination of errors and failures by strict application of the rules and by the control of that application.
6) Safety also depends on the capacities of actors and organizations to learn from incidents, dysfunctions, quasi-accidents and accidents which, to varying degrees, might have undermined security.

These different assumptions all warrant closer examination.

The absolute priority given to people's safety is a form of compensation for the inevitable development of risks in modern societies, due to technical progress. It relates to considerations of a moral order, to the necessary preservation of the general interest, to the future of human communities, etc. This priority determines the obligations and responsibilities of the authorities and organizations directly in charge of dangerous activities and those which, in state administrations, define public policies or occupy functions of control and expertise. The attribution of these responsibilities requires skills (technical, scientific, organizational, etc.) in those who create and manage risks and carry out controls. Moreover, it implies an 'enlightened confidence', even an enlightened consent, as regards these managers, especially the public authorities responsible for guaranteeing safety.

The recommendation of maximum efforts (in terms of knowledge and action) derives from the absolute priority given to the security of persons. The calculation of the costs of security is consequently played down; it is assumed that they can always be absorbed, considering the possible consequences of the materialization of certain risks. This priority is based on the possibility of an effective risk control through the mobilization of science and techniques, and the development of effective action linked to permanent vigilance among the authorities and organizations in charge of risks. More recently, it has been based on the application of the precautionary principle when scientific and technical knowledge does not yet allow the risk to be defined with precision.

As regards knowledge, it is possible to model and control the uncertainty associated with risk: identify the danger, define an acceptable risk, model the processes involved, measure or evaluate potential damage and its probability, in the present or the future, define strategies and methods of risk reduction (e.g., reduction of the danger – risk at the source – of the probability of damage, of the damage itself, etc.) and, finally, define criteria of priority and decisions concerning risk.

Prior knowledge of risks makes prevention, recovery and mitigation possible. Three risk-reduction strategies predominate: first, eliminating the risk at its source; second, quality and the monitoring of sure procedures – a guarantee of end performance and safety; and third, construction of in-depth defences or barriers against the risk. Risk control relates essentially to a normative strategy consisting of the modelling, description and specification of an operationally perfect world. It involves the prescription of actions and behaviours which are supposed to ensure that one remains within that world (good design rules, operational procedures), and the deployment of actions to prevent or harness the feared events, and to attenuate their consequences. It is based on approaches such as Total Quality or its modern, more realistic versions (Continuous Quality Improvement). Finally, risk control aims progressively to establish a system of prescriptions that are as perfect as possible, and to obtain the closest possible adhesion of involved actors to this system of prescriptions.

Safety stems primarily from compliance with all the norms, rules and procedures. This is the objective of the effort to produce knowledge and actions designed to frame dangerous activities. Non-compliance with these rules threatens security, to varying degrees, whether that non-compliance is the result of errors (non-intentional behavior) or violations (corresponding to obvious and intentional infringement of the rules). Safety actions and policies therefore focus on the avoidance of errors and violations, and especially on the frontline actors (the ‘operators’) who intervene at the end of the chain. Even if it is only one of the possible means of intervention, the principle of punishing mistakes, that is, errors and violations judged unacceptable (because considered to be avoidable by an average professional and having had – or which could have had – serious consequences for security) is affirmed. Above all, violations of law, and the liberty taken in this respect, are punishable. A certain margin for adjustment in the interpretation of rules is sometimes tolerable, but cannot under any circumstances justify systematic slackness and risk-taking.

Security is also an outcome of feedback, which seems to be the most obvious and the most appropriate practice for analysing the occurrence of errors and failures – especially in relation to departure from the norms, rules and procedures, in other words, the overstepping or circumvention of the various defence barriers. Feedback is therefore conceived of in relation to the devices that are provided for to frame high-risk activities. It is intended primarily to identify and explain departures (from the reference framework) and to reduce them to residual deviations.

The dominant safety paradigm in these six assumptions is therefore organized essentially around the ideas of responsibility (whether claimed or assigned), the assertion of the power of science and techniques, and the effectiveness (even the virtue) of the rule in the quasi-legal sense of the term.

But this reassuring and rational discourse fools no one. All the concrete actions are matched with another discourse introducing a series of corrections and adjustments in the name of the principle of reality. For instance, it is agreed that in the world of industry and transport the priority given to risks cannot be absolute (considering the other imperatives), that the knowledge of risks may be imperfect (especially due to the complexity of the activities and processes, and the irreducible part of certain uncertainties), and that the rules designed to frame action may be insufficient, not entirely appropriate, not fully complied with, and so on. Likewise, safety is not considered as a result only of the application of rules and the implementation of feedback procedures; it also depends on the degree of development of a 'culture of security', of the establishment of 'trust' between actors, etc.

These corrections are made, however, without any real questioning of the underpinnings of this paradigm. Although they are recognized as being difficult to attain, 'zero risk' and 'total quality' are still held up as desirable and even possible objectives. But, faced with such demands, which relate to an essentially normative risk management without concessions, the various trade-offs and compromises running through risk management cannot be reduced simply to corrections and adjustments. The risk versus economic performance compromise is thus discussed daily and precludes any achievement of this security ideal.

The security model tacitly adopted, based on an ideal reference framework (zero risk, total quality), is therefore unrealistic. But as a slogan, a belief, it is easy to understand and serves as a motivation for all and a reassurance for the general public and – sometimes – for its leaders. And when strong constraints generated by the dominant paradigm make it impossible not to acknowledge a possible loss of control (as regards knowledge, vigilance, action, etc.), a sliding towards a substitute approach based on caution and precaution rapidly tends to occur. These changes correspond to a significant shift of the safety issue since the main objectives are the cancellation of the risk or its mitigation (imputing liabilities, paying out compensation). Currently there is either an alignment of discourses and analyses on the dominant paradigm (with the introduction of a moderate version which does not challenge it) or else an exit from this paradigm on the basis of caution and precaution (without other approaches being considered).

First Critique of the Dominant Paradigm: No World is Perfect

A systematic critique of the assumptions underpinning this dominant paradigm led us to raise serious doubts about the paradigm itself. For each of the assumptions, strong elements have been put forward, so that the entire paradigm can be challenged. This point-by-point deconstruction is based on various sources: the international literature, especially studies exploring obtaining safety via other ways than control and rules; French research studies, especially in sociology, political science and economics, which account for multiple constraints weighing on the definition and implementation of security policies; and, finally, research in ergonomics and cognitive psychology, which have provided the main lever for organizing a critique of the dominant paradigm.

For the organizations managing high-risk activities, people's safety is only a priority among others (assumption 1). Performance, competitiveness, the continuation of activities, the preservation of capacities for technological innovation, experimentation, etc. are other imperatives for these organizations faced with the constraints of the market or of the public service. To a large extent, the agenda of these organizations is also shaped by deregulation policies, accelerated technological change, flexibility challenges, etc. (Rasmussen, 2001). Even if it is displayed as an absolute priority, safety necessarily involves variable trade-offs, depending on the type of organization and activity. These trade-offs weigh on the modes of structuring and functioning of organizations. Thus, the concept of organizational reliability has to combine economic efficiency and safety (Bourrier and Laroche, 2001). Moreover, safety does not amount to the safety of humans, productive systems, goods and the environment. It also needs to be understood in relation to other imperatives, for instance the permanence of activities resulting from socio-technical systems' tolerance to design errors, malicious attacks (Laprie, 2001), or changes in organizational structures and processes, among others.

The efforts to maintain or enhance safety within organizations managing high-risk activities are necessarily limited by the resources (in the broad sense of the term) assigned to safety and therefore by the explicit or implicit trade-offs between the various imperatives, both within the organizations responsible for these activities and between them and the control authorities or the expertise agencies (assumption 2). These trade-offs determine which share of the means, competencies and legitimacy is assigned to safety actions and policies. Other limits are related to the building of commitment around safety policies. When they are labelled as such, safety policies are difficult to promote, to implement and to embed lastingly in organizations' functioning. They often require a strong will in the actors and institutions in charge of them. They may seem excessively restrictive or even inappropriate compared to the problems they are supposed to solve (for instance, due to their high level of specificity). It may also be considered that overall safety depends only partially on actions and policies narrowly defined as safety-related (safety can, for example, be dissolved in quality and taken into account in a routine mode rather than in one of mobilization).

Organizations' 'knowledge of risks' is also limited in various ways (assumption 3). Organizations, which do not know themselves well

(Rochlin, 2001; Rasmussen, 2001), use only a part of the knowledge that could be mobilized. Depending on the type of organization and activity, choices are made among the available data (factors, events, etc.) and risk models. These choices are conditioned by the intensity of the research effort – which itself is highly variable – by the amplitude of feedback procedures – which depend on organizations' capacity to carry the 'costs' (financial, organizational, social, even political) related to these procedures (de Keyser, 2003; Gilbert, 2001) – and by shared experiences within and between organizations. They vary, depending on the modes and levels of technical, scientific, organizational and other investments made, and are strongly shaped by what has already been tried and tested. As in all organizations faced with different types of constraints and objectives, the usual tendency is to repeat known cognitive approaches and thus to see problems as they have already been defined and solutions as they have already been devised in the past (Bourrier and Laroche, 2001). Reproducing safety policies and behaviours makes it possible to stabilize knowledge and reference frameworks at the lowest cost (technical, cognitive, social, political, etc.). But it can also make it difficult to question and to challenge them (especially if the level of safety supposedly obtained is considered as acceptable). Incorporating new factors into current procedures (the organizational factor, for example) can thus be seriously hindered. Consequently, ballistic approaches of risk – prior definition of risks, determination of a mode of framing activities, measurement of deviations – tend to dominate over more dynamic ones – as translated in the concept of loss of control (Paries, 2001, 2003) or in the idea of the management of safety margins (Amalberti and Malaterre, 2001).

Safety actions also encounter different types of limits. The translation of knowledge on risks into action is not automatic; it depends on the knowledge drawn upon and on the constraints related to that, but equally on the types of action that the organization can effectively undertake and the costs that it can carry (assumptions 4 and 6). The production of norms, procedures and rules, the creation of data bases, etc. are therefore given priority. Accordingly, new knowledge is not systematically integrated – for instance knowledge on dynamic monitoring of risks (Hollnagel, 2001), or on switching from a type of management based on experience to one based on vigilance, anticipation and integration of the future into the present (Clot, 2002). The same applies to those aspects and problems which are easiest to grasp (e.g., technical aspects or limited events without serious consequences), rather than those which may appear to be the most urgent to treat (organizational aspects (Bourrier and Laroche, 2001); major accidents, crisis situations (Lagadec, 2002)). Moreover, safety actions are certainly defined in relation to the problems to solve and the knowledge and know-how mobilized, but also in relation to other criteria such as compliance with the rules, the prior definition and allocation of responsibility, etc.

Limited in various ways, safety actions are also considered to contribute only partially to safety in high-risk activities. The idea of maximum framing of these activities, as conveyed by these actions, is replaced de facto by more relative and pragmatic conceptions which appear to be the real foundations of safety in organizations. The rules (broadly-speaking) produced by safety policies and actions are displayed and constitute key references, but they neither define nor determine the practices of the different actors within those organizations (assumption 5). The actors, irrespective of their level, although this is particularly true for those involved in running and monitoring high-risk activities, base their practices on the search for the best possible control of such activities by incorporating various types of constraints (general as well as local) (Clot, 2002; Fleury, 2002; de Keyser, 2003). They place themselves in the frame of sufficiency (Amalberti and Malaterre, 2001) rather than striving to satisfy the demands related to norms, rules, procedures and so on. Conversely, they are not systematically trying to transgress them (Bourrier and Laroche, 2001). They position themselves in more intermediate zones: the operator's 'improvisation', for instance, can be seen as being regulated by a system of 'informal' norms at the intersection of top-down and bottom-up rules. Likewise, they do not endeavour to mobilize all available cognitive and action resources (Thoenig, 1995), to change their 'cognitive frame' and to open up to new 'registers of action'; still, they are not closing up to vigilance, attention and different forms of anticipation (Clot, 2002). Above all, in their environment filled with constraints, they seek to obtain at a low cost (in terms of energy, attention, etc.) 'sufficient' capacities and resources to treat the problems they have to solve. Thus, at a certain level the cognitive processes and human actions constantly deviate from any external standard. They aim for neither an optimum nor perfection; they are only 'sufficient' to attain the goal. Thus, a mode of functioning of normally 'sub-optimal' organizations emerges, normally out of line with the injunctions of safety policies and actions.

Second Critique of the Dominant Paradigm: What is More Natural Than Errors and Failures?

A second, more precise critique concerns the very foundations of the dominant paradigm by questioning the status of an 'error'. This critical approach is based on overlapping evidence gathered from various disciplines.

From a neurophysiological point of view (Hasbrouck *et al.*, 2001, 2002; Falkenstein, 2001), the nervous system does not 'try to' work without errors. In simple tasks it has been proved that it prepares all the possible answers, both good and bad, and often simultaneously launches a faulty execution and an error-detection loop. The wrong response is then inhibited and the right one facilitated. Nothing is visible from the outside apart from a slightly longer response time. In short, the neural architecture functions more on the basis of control and detection-recovery than on a perfect faultless production. The gain is probably considerable as regards response speed and adjustment.

From a macro cognitive point of view (Amalberti and Malaterre, 2001; Hollnagel, 2001), the picture is more or less the same. Error is inseparable from human intelligence; it is in a sense 'the other side of the coin' (Reason, 1990). This is nothing new but the explanatory models are. In fact, error is not attached to a human operator like a natural punishment for his or her limitations; on the contrary, it reflects the result of constant cognitive choices related to strategies that humans adopt to by-pass their intellectual limitations and to find the best compromise between performance and acceptable risk (or perceived as such), in view of their own limitations/ constraints and those of the environment and the organization.

The sociological approach to organizational management, just as the preceding levels of analysis, suggests that vulnerabilities are part of the natural logics of corporate life. The difficulties encountered in correctly processing information are the result of the same mechanisms that allow its processing (established cognitive frames such as worldviews, ideologies or identity filters). They expose the organization to a risk of cognitive closure, that is, a gradual blindness and functional rigidity. Conflicts of interest between production (efficiency, speed, reactivity, comfort) and safety naturally lead to deviance from the safety reference framework, which can be normalized in the long term and result in an accident.

This reflection also draws on studies which distinguished between mistake and error (Amalberti and Malaterre, 2001). In 1981 Norman introduced a fundamental distinction between mistakes (errors of intention and conscious processing) and slips (errors of execution, and automatic processing). More recently, Leplat (1985), Rasmussen (1986), Senders and Moray (1991) and above all Reason (1993) have proposed a series of models on the mechanisms of error and the circumstances in which they occur. To this they have added and explained the particular category of intentional 'errors', still called violations. Many studies on the subject in the French-speaking community are grounded in this heritage, for example: in Liège, Keyser and Nyssen (2001) (errors related to time), Masson (1994) (errors in routines), Nyssen (2000), Nyssen and De Keyser (2001) (errors in medicine); in Paris, Amalberti (1996), Amalberti and Pariès (2000), Pariès (1994, 1995) (errors in aeronautics), Wioland (1997) (detection of errors).

This reflection also draws on studies which have situated error in an accidental process (Amalberti and Malaterre, 2001). Two dominant features emerged from this literature from the 1980s and have had numerous consequences on safety: 1) error is inseparable from human intelligence (as already stated above); 2) accidents are not (directly) related to operators' errors. Rather, they are related to the situations in which human error may have occurred and, following a 'chain of events', have led to an accident due to a lack of protection in the system. In short, an accident is simply the evidence of one or more bad defences in the system as a whole. A sure system has to allow errors or failures and protect itself against the consequences of such events. This is a systemic point of view. To develop it further, Reason

(1990) introduced two important ideas: the distinction between latent error and patent error, and the concept of in-depth defence.

Taking into account these advances leads to a very different approach to error which is still largely taboo in risk research. It has nevertheless been explored, notably through research in cognitive ergonomics and psycho-ergonomics (Amalberti, 1996, 1998). Most of these studies concern individuals ('operators') or small collectives in specific fields (fighter planes, commercial planes, trains, boats, etc.) in which advanced technology is used.

Observations of operators' practices show that they regularly make errors, which are therefore not abnormalities or exceptions. The errors are accepted, remedied or ignored. Errors are a price to pay, a necessity for adjustment, mere symptoms of good cognitive functioning. The only risk for the operator is to lose control of the situation and not to do the job. This is the risk that the operator really manages (Amalberti and Barriquault, 1999). Strong hypotheses are based on this reasoning: the occurrence of errors and the fact of them being forgotten are not necessarily the sign of a lack of safety if they are accompanied by a focus on the factors or processes that allows the situation to be kept under control; the occurrence of errors *and* their being remedied can be a factor of safety since the control of a process often seems to be related to this dynamic. It is noted that accidents and disasters often occur when there is a deregulation of the risk model associated with different errors, an excessive focus on the elimination of all possible errors when the errors occurrence/detection/recovery sessions are no longer present or are no longer made visible, due to the technologies implemented – mainly automation with 'black box' effects.

Via the recognition of the habitual non-optimality, sufficiency and normality of the error forgotten or recovered, one can throw the bases of a new approach to safety – *'ecological safety'* (Amalberti, 1996) – in which a number of taboos are questioned. Founded on a delimited empirical base (operators, small collectives), this approach opens onto more general reflection. It leads to questions on *'ultra-sure'* systems such as those prevalent in the nuclear field and aviation (Amalberti, 1996). Here accidents are now too rare for a connection to be made with data on problems that arise ordinarily. Moreover, the automation of processes and of corrections of human errors and other dysfunctions is such that possibilities of adjustments seem to be limited. Finally, the changes made to increase reliability defy the monitoring capabilities of the actors in charge of these high-risk activities within the organizations. In other words, the issue of the loss of control in certain types of activity has to be raised again. This loss of control seems to be related to a progressive opacity of systems which, even though they appear to be extremely sure, are potentially highly vulnerable due to the low level of readability of their actual mode of functioning and dysfunctioning.

Errors (and all failures) can neither be reduced to departures from the rules, nor considered as abnormalities or exceptions. They are an integral part of habitual, normal functioning, irrespective of the level on which they are situated. As such, they have an ambivalent status from the safety point of view. On the one hand the persistence in error or failure without correction or remedies, the habit of deviating (the *'normalization of deviance'* (Vaughan, 1996, 2000)), can lead to serious dysfunctions, accidents and even disasters. On the other, errors and failures which are repaired not only maintain high-risk activities within acceptable bounds of safety, they also allow an effective risk control (via the experience of errors and deviances). Hence, the focal point for safety is shifted. The dominant 'normative' strategy is replaced by a 'natural' or 'adaptive' strategy, so that risk management is based not on a striving for 'perfection from the start' nor even on systematic surveillance of 'deviances' for the purpose of correcting them, but on constant surveillance of the safety margins and levels of risks taken. In other words, safety is based on a 'controlled approximation' rather than on a search for the minimum departure from a reference framework. Metaphorically, this is an immunity type of process in which defences are built and feed on vulnerabilities. One implication is that, if threats disappears or are masked in some way, defences cannot develop. In this perspective, the challenges lie in the conception of error-tolerant systems (following research on safety in the functioning of computer technology (Laprie, 2001)) and on the building of an 'organizational robustness' (following research in the sociology of labour, especially concerning the structuring of work collectives (de Terssac, 2001; Boissières and de Terssac, 1999a, 1999b; Boissières and Marsden, 2005)).

Beyond the Dominant Paradigm?

Critique of the dominant paradigm is accompanied by the emergence of a new 'safety paradigm' based on neither injunctions nor necessities but rather on what usually happens in the normal course of high-risk activities. In contrast to an ideal strategy aimed at the total control of risks, an alternative is proposed, corresponding to the modelling of empirical findings on risk management and the status of errors. A new set of assumptions thus emerges, the main three of which would be the following:

- The first is the recognition of the various *compromises* that the actors and organizations involved to varying degrees in the management of high-risk activities have to reach. From organizations in charge of a high-risk activity down to individual operators, the combination of multiple constraints and imperatives induces compromises. Reflection has then to focus on the relationship between compromise and safety, on the terms of negotiation of such compromises, on the processes by which they are reached, maintained or challenged, on the establishment of criteria for distinguishing good, less good, and bad compromises, etc.
- The second is the recognition of the *normal character of failures* in high-risk activities. Technical failures, human errors and organizational dysfunctions reflect the naturally unstable nature of these activities, the

stability of which is consequently based on continuous adjustments and reparations. Reflection has then to focus on the conditions in which an acceptable state of safety is maintained (considering this initial instability and the dynamics that it generates), and on the tolerance to different types of failure.

- The third is the recognition of the *normally sufficient, non-optimal* nature of actors' practices, and more generally of the functioning of organizations or networks of organizations involved in risk management. This sufficiency and non-optimality appear to be the required conditions for maintaining a control and steering capacity in activities characterized by compromise and instability. Reflection has then to focus on the modes of emergence of the practices of sufficient, non-optimal functioning and on the way of conceiving of new forms of risk management.

On the basis of these assumptions, it seems to be possible to go beyond the observation of obvious differences between the dominant safety paradigm and current practice. It also appears to be possible not to limit ourselves to the solutions usually recommended for reducing those differences, whether these concern the implementation of feedback procedures, the development of a safety culture within firms and organizations, or the generalization of relations of trust, for example. Significant changes or even total departure from former views of safety therefore seem conceivable, both conceptually and in practice, when the approach to safety issues is based on actual practices and modes of functioning. Yet, unsurprisingly, various obstacles hinder such changes of approach.

First, it is obvious that this reflection is not mature enough. Efforts are still needed to define the new concepts which would enable us to flesh out the new paradigm, other than in relative terms. The notions of '*compromise*', '*normal error*', '*sufficiency*' and '*non-optimality*' are more a matter of markers of a critique than a real change of paradigm. Hence, the reflection underway has not yet been able to break away from critique of the dominant paradigm which still serves as a reference. Moreover, the assumptions based on the propositions of psycho-ergonomics and cognitive ergonomics, and enhanced by many other studies and perspectives, have provided a framework of analysis that is still weak from a theoretical point of view. This is due primarily to the diversity of approaches and the difficulty sometimes encountered in going beyond similar identifications and reasoning of a metaphorical nature. For example, the idea of an alternative so-called 'adaptive' approach to risk-management, derived from the modelling of individual cognitive processes, is still a hypothesis waiting to be formalized, consolidated and validated, especially at the higher organizational levels (team, organizations, society). More generally, to support these new assumptions it still seems necessary to mobilize an abundant existing literature which has not yet been analysed in these terms and the conclusions of which remain scattered. Studies concerning the definition of problems, framing and agenda-setting processes, for example, can help to understand how compromises between different types of constraints are reached within organizations and groups of organizations. Work on 'ordinary pathologies' of actors and organizations is also valuable for analysing the reasons for which *'sufficiency'* is often preferred to *'optimality'*, and error or failure becomes *'normal'* (Laroche, 1996). It is likewise necessary to examine, from these perspectives, work aimed at determining the conditions of organizational reliability and, more particularly, that relative to organizations considered to be sure or even ultra-sure (*High Reliability Organizations*) (Bourrier, 2001). Based on the approach of the Berkeley group (La Porte, 1994, 1996, 2001; La Porte and Consolini, 1991; Rochlin *et al.*, 1987; Rochlin, 1991, 1996; Schulman, 1996) and then Sagan (1993) and Heimann (1993, 1997), the actual motivations of that reliability are taken into account (apart from the systems and rules established). Likewise but in a different way, the work of Karl Weick around the notion of 'sensemaking' (1995) analyses the mechanisms allowing human collectives to withstand critical events and situations (Laroche, 1996; Koenig *et al.*, 2003). Thus, if we want to switch from a critique of the dominant paradigm to the elaboration of a new paradigm, we need to reconsider the results of a large number of existing studies from a different perspective.

The conception of new tools related to the new 'paradigm' has also appeared to be fairly problematic. Among issues that remain problematic, one can mention, for example: proving that safety is obtained by mechanisms other than the definition of and adherence to a framework of action, mainly by highlighting a statistical de-correlation between the frequency of events representative of deviations on a low level (errors, violations, incidents) and the frequency of dreaded events (serious accidents and disasters); modelling processes of perception of the 'safety space' and permanent markers or processes of risk-evaluation associated with the variations, errors and failures and, more generally, modelling processes of control and loss of control, of stabilization and destabilization of self-organized systems; determining whether it is possible to 'deliberately predict the exit from the norm' and to 'frame' that exit, etc. But these perspectives, which are embedded in the project to dynamically monitor activities and thereby depart from the 'ballistic perspective', do not appear to lead (at least in the short term) to the conception of tools allowing us to define fairly simple management methods (unlike the approaches where the output is the definition of norms, rules and procedures). One of the questions which arise concerns the possibility not only of conceiving of new tools but also of being able to experiment with them. Currently there are still few of these in France (for example, the PREDIT/NAOS project of the SNCF) and in Northern Europe. Another difficulty clearly pointed out is that these new conceptions presuppose attention and vigilance which, in various respects, may seem costly and difficult to maintain in the long term. Unlike the 'ballistic' approach, the 'adaptive' approach probably demands high investments, constant attention by all the actors at all levels of the organization, and a high level of latent reactivity.

But it is above all at a socio-political and political level that the main obstacles appear. The view of high-risk activities and their management via compromise, normal error and failure, sufficiency and non-optimality, obviously poses problems of acceptability and receivability, given the nature of current debate on risk. Even though the actors directly involved in the management of dangerous activities often acknowledge the necessity to change the paradigm and the reference framework, they usually highlight the 'risks' of such a change – especially since significant progress would need to be made from a conceptual and operational point of view, as well as in the redefinition of responsibilities, if this type of project were launched (unless experimentally, as proposed). The extent of the changes to make has not been underestimated; but, despite their justification in paradoxical terms of 'compromise' and 'sufficiency', these changes are necessary for effectively improving safety in the future. Likewise, the difficulties have been clearly identified, especially among decision-makers. Even if the direct managers of risks can recognize the actual reality of risk management and even promote that recognition, its acceptance will remain problematical for public authorities and leaders. One of the obstacles identified is the difficulty of reverting to the 'tacit contract' between the different stakeholders in safety matters. In France, at least, that contract is based on a delegation to the State: the latter is expected to provide for safety and collective security (with the implied consequences in the attribution of responsibility, competence, and trust). Questions of power and legitimacy associated with these themes are such that it still seems difficult to be able to really analyse the issue of 'technical democracy', that is, to envisage a collective definition of acceptable risk, a relative distribution and weighing up of the pros and cons, as well as a definition and verification of the competencies of risk managers.

One of the conclusions of interdisciplinary reflection engaged around questions of error and failure in contemporary socio-technical systems is thus that one of the first deadlocks in the development of new conceptions concerning safety is of a political order. To be sure, the difficulties, both conceptual and operational, are great. Only the establishment of a large-scale research programme, encompassing diverse disciplines, would allow the announced change of paradigm to happen. Likewise, only the multiplication of experiments would make it possible to test new tools, new procedures, coming from fundamental and applied research and expertise. But launching a project such as this means that a set of socio-political conditions have to be met – and this relates to policies or, more exactly, to political debate on safety. We wonder whether this debate can take place in modern democracies, and especially in France. One of the most salient questions is whether the incumbent authorities still hold that they are the only ones authorized to define approaches and policies in this domain or whether, on the contrary, they admit that they can henceforth put safety issues to debate and thereby possibly help to bridge the gap between discourse on risk-control and actual risk-management. This would mean that, collectively, our societies agreed not to lie to themselves about safety issues.

References

Amalberti, R. (1996) *La conduite de systèmes à risques*, (Paris: PUF) [2nd edition, 2001].

Amalberti, R. (1998) Notions de sécurité écologique: le contrôle du risque par l'individu et l'analyse des menaces qui pèsent sur ce contrôle. Approche psycho-ergonomique, in: *Actes de la 9^{e} séance du Séminaire du Programme Risques Collectifs et Situations de Crise* (Grenoble: CNRS).

Amalberti, R. and Barriquault, C. (1999) Fondements et limites du retour d'expérience, *Annales des Ponts et Chaussées*, 91, pp. 67–75.

Amalberti, R. and Malaterre, G. (2001) De l'erreur humaine au risque: évolution des concepts en psycho-ergonomie, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Risques, erreurs et défaillances. Approche interdisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 71–106.

Boissières, I. and de Terssac, G. (1999) Organizational conflicts in safety interventions, 17th International Workshop "New Technologies and Work", Bad Homburg, Germany, 17–19 June.

Boissières, I. and Marsden, E. (2005) Organizational factors of robustness, in: B. Carle & B. Van de Walle (Eds) *Proceedings of the Second International ISCRAM Conference* (Brussels, Belgium: ISCRAM).

Bourrier, M. (2000) *Théories et pratiques de la fiabilité organisationnelle, Rapport final pour le Programme Risques Collectifs et Situations de Crise du CNRS, COSTECH*, Compiègne: Université de Technologie de Compiègne.

Bourrier, M. (Ed) (2001) *Organiser la fiabilité* (Paris: L'Harmattan).

Bourrier, M. and Laroche, H. (2001) Risque de défaillance : les approches organisationnelles, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Risques, erreurs, défaillances. Approche interdisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 15–51.

Clot, Y. (2002) La place des autres dans le travail du conducteur de train, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Conditions et mécanismes de production des défaillances, accidents et crises* (Grenoble: CNRS-MSH-Alpes). pp. 293–312.

De Keyser, V. and Nyssen, A. S. (2001) The management of temporal constraints in naturalistic decision making. The case of anaesthesia, in: E. Salas & G. Klein (Eds) *Linking Expertise and Naturalistic Decision Making* (Mahwah, NJ: Lawrence Erlbaum). pp. 171–188.

De Keyser, V. (2003) Les systèmes de report d'incidents, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Autour de la mesure du risque. Un questionnement multidisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 41–72.

Fleury, D. (2003) Conditions de survenue des accidents graves de la route et du travail. Les accidents de la route et leur prevention, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Conditions et mécanismes de production des défaillances, accidents et crises* (Grenoble: CNRS-MSH-Alpes). pp. 91–110.

Falkenstein, M. (2001) Action errors and brain activity, Paper delivered at the 4th session of the seminar "Le risque de défaillance et son contrôle par les individus et les organisations dans les activités à hauts risques", CNRS (RCSC) - Ministère de la Recherche (ACI Cognitique), Gif-sur-Yvette, France, 14–15 May.

Gilbert, C. (2001) Retours d'expérience : le poids des contraintes, *Annales des Mines. Responsabilité et Environnement*, 22, pp. 9–24.

Girin, J. and Grosjean, M. (Eds) (1996) *La transgression des règles au travail* (Paris: L'Harmattan).

Hasbroucq, T., Burle, B., Bonnet, M., Possamaï, C. A. and Vidal, F. (2001) Arguments physiologiques en faveur d'un contrôle d'exécution au cours d'activités sensorimotrices sous contrainte temporelle, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Risques, erreurs et défaillances. Approche interdisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 239–260.

Heimann, C. F. L. (1993) Understanding the Challenger disaster: organizational structure and the design of reliable systems, *American Political Science Review*, 87, pp. 421–425.

Heimann, C. F. L. (1997) *Acceptable Risks, Politics, Policy, and Risky Technologies* (Ann Arbor, MI: The University of Michigan Press).

Hollnagel, E. (2001) Accident modelling and performance variability management, Paper delivered at the 4th session of the seminar "Le risque de défaillance et son contrôle..." CNRS (RCSC) - Ministère de la Recherche (ACI Cognitique), Gif-sur-Yvette, France, 14–15 May.

Koenig, G., Allard-Poesi, F., Vidaillet, B., Laroche, H. and Roux-Dufort, C. (2003) *Le sens de l'action. Karl Weick : sociopsychologie de l'organisation* (Paris: Vuibert).

Lagadec, P. (1981) *La civilisation du risque. Catastrophes technologiques et responsabilité sociale* (Paris: Seuil).

Lagadec, P. (1988) *Etats d'urgence. Défaillances technologiques et déstabilisations sociales* (Paris: Seuil).

Lagadec, P. (with Guihou, X). (2002) Les conditions de survenue des crises graves, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Conditions et mécanismes de production des défaillances, accidents et crises* (Grenoble: CNRS-MSH-Alpes). pp. 157–210.

La Porte, T. (1994) A strawman speaks up: comments on the limits of safety, *Journal of Contingencies and Crisis Management*, 2(4), pp. 207–11.

La Porte, T. (1996) High reliability organizations: unlikely, demanding and at risk, *Journal of Contingencies and Crisis Management*, 4(2), pp. 60–71.

La Porte, T. and Consolini, P. (1991) Working in practice but not in theory: theoretical challenges of 'High Reliability Organizations', *Journal of Public Administration Research and Theory*, 1, pp. 19–47.

La Porte, T. (2001) Fiabilité et légitimité soutenable, in: M. Bourrier (Ed.) *Organiser la fiabilité* (Paris: L'Harmattan). pp. 71–106.

Leplat, J. (1985) *Erreur humaine, fiabilité humaine dans le travail* (Paris: A. Colin).

Laprie, J.-C. (2001) Sûreté de fonctionnement informatique, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Risques, erreurs et défaillances. Approche interdisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 123–147.

Laroche, H. (1996) Risques, crises et problématique de la décision. Point de vue de Hervé Laroche, in: *Actes de la 4^{e} séance du Séminaire du Programme Risques Collectifs et Situations de Crise du CNRS* (Grenoble: CNRS).

Masson, M. (1994) Prévention automatique des erreurs de routine. Thèse de Doctorat en Psychologie, Université de Liège.

Norman, D. (1981) Categorization of action slips, *Psychological Review*, 88, pp. 1–15.

Nyssen, A. S. (2000) Analysis of human errors in anaesthesia. Our methodological approach: from general observations to targeted studies in laboratory, in: C. Vincent & B. A. De Moln (Eds) *Safety in Medicine* (London: Pergamon). pp. 49–63.

Nyssen, A. S. and De Keyser, V. (2001) Prevention of human errors in the frame of the activity theory, *International Handbook of Work and Organizational Psychology*, 1(10), pp. 348–363.

Pariès, J. (1994) Etiology of an accident: human factors aspects of the Mont Sainte-Odile crash, *ICAO Journal*, 49, (6).

Pariès, J. (1995) Evolution of the aviation safety paradigm: towards systemic causality and proactive actions, Paper presented at the Third Australian Aviation Psychology Symposium, Sydney, Australia.

Pariès, J. and Amalberti, R. (2000) Aviation safety paradigms and training implications, in: N. Sarter & R. Amalberti (Eds) *Cognitive Engineering in Aviation* (Hillsdale, NJ: Lawrence Erlbaum Associates). pp. 253–286.

Pariès, J. (2003) Résumé de l'épisode précédent : retour sur les définitions du risqué, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Autour de la mesure du risque. Un questionnement multidisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 9–18.

Perrow, C. (1984) *Normal Accidents, Living with High-Risks Technologies* (New York: Basic Books).

Perrow, C. (1999) Organisations à hauts risques et normal accidents, *Actes de la 14e séance du Séminaire du Programme Risques Collectifs et Situations de Crise* (Grenoble: CNRS).

Posamaï, C. A., Burle, B., Vidal, F. and Hasbroucq, T. (2002) Conditions de survenue des défaillances dans les tâches sensorimotrices, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Conditions et mécanismes de production des défaillances, accidents et crises* (Grenoble: CNRS-MSH-Alpes). pp. 225–248.

Rasmussen, J. (2001) Accident causation and risk management: basic research problems in a dynamic, tightly coupled society, Paper delivered at the 4th session of the seminar "Le risque de défaillance et son contrôle..." CNRS (RCSC) - Ministère de la Recherche (ACI Cognitique), Gif-sur-Yvette, France, 14–15 May.

Reason, J. (1990) *Human Error* (Cambridge: Cambridge University Press).

Rochlin, G. (1991) Iran Air Flight 655 and the USS Vincennes: complex, large-scale military systems and the failure of control, in: T. La Porte (Ed.) *Social Responses to Large Technical Systems: Control or Anticipation* (Norwell: Kluwer Academic Publishers). pp. 95–121.

Rochlin, G. (1996) Reliable organizations: present research and future directions, *Journal of Contingencies and Crisis Management*, 4, pp. 55–60.

Rochlin, G. (2001) Nobody sees the trouble I've known: risk and safety in complex technical systems, Paper delivered at the 4th session of the seminar "Le risque de défaillance et son contrôle..." CNRS (RCSC) - Ministère de la Recherche (ACI Cognitique), Gif-sur-Yvette, France, 14–15 May.

Sagan, S. (1993) *The Limits of Safety: Organizations, Accidents and Nuclear Weapons* (Princeton, NJ: Princeton University Press).

Senders, N. and Moray, J. (1991) *Human Error: Cause, Prediction and Reduction* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Schulman, P. (1996) Heroes, organizations and high reliability, *Journal of Contingencies and Crisis Management*, 4, pp. 72–83.

Setbon, M. (1995) L'action organisée en réponse au risque-sida tranfusionnel, in: *Actes de la 2e séance du Séminaire du Programme Risques Collectifs et Situations de Crise du CNRS* (Grenoble: CNRS).

de Terssac, G. (2001) Les risques de la rationalisation du point de vue de la sociologie du travail, in: R. Amalberti, C. Fuchs & C. Gilbert (Eds) *Risques, erreurs et défaillances. Approche interdisciplinaire* (Grenoble: CNRS-MSH-Alpes). pp. 169–194.

Thoenig, J.-C. (1995) L'action collective organisée face au risque: d'un cadre conceptuel au cas du risque-sida, in: *Actes de la 2e séance du Séminaire du Programme Risques Collectifs et Situations de Crise du CNRS* (Grenoble: CNRS).

Vaughan, D. (1996) *The Challenger Launch Decision. Risky Technology, Culture, and Deviance at NASA* (Chicago, IL: University of Chicago Press).

Vaughan, D. (2000) Technologie à hauts risques, organisation, culture. Le cas de Challenger, in: *Actes de la 15e séance du Séminaire du Programme RCSC* (Grenoble: CNRS).

Weick, K. (1987) Organizational culture as a source of high reliability, *California Management Review*, 29, pp. 112–127.

Weick, K. (1993) The collapse of sensemaking in organizations: the Mann Gultch disaster, *Administrative Science Quarterly*, 38, pp. 628–652.

Wioland, L. (1997) Etude des mécanismes de protection et de détection des erreurs, contribution à un modèle de sécurité écologique, PhD thesis in the psychology of cognitive processes, Université Paris V.