

The Fallacy of Severity Classification in Risk Assessment Methods

Alberto Pasquini¹, Simone Pozzi^{1 2} and Luca Save^{1 3}

¹Deep Blue srl, Rome, Italy

²Sapienza University of Rome, Department of Psychology of Social and Developmental Processes, Rome, Italy

³University of Siena, Media and Communication Department, Siena, Italy
{alberto.pasquini; simone.pozzi; luca.save}@dblue.it

Abstract

The knowledge of operational experts plays a fundamental role in performing safety assessments in safety critical organizations. The complexity and socio-technical nature of such systems produce hazardous situations which require a thorough understanding of concrete operational scenarios and cannot be anticipated by simply analyzing single failures of specific functions. This paper addresses some limitations regarding state-of-the-art safety assessment techniques, with special reference to the use of severity classes associated to specific outcomes (e.g. accident, incident, no safety effect, etc.). Such classes tend to assume a linear link between single hazards considered in isolation and specified consequences for safety, thus neglecting the intrinsic complexity of the systems under analysis and reducing the opportunities for an effective involvement of operational experts. An alternative approach is proposed to overcome these limitations, by allowing operational people to prioritize the severity of hazards observed in concrete operational scenarios and by involving them in the definition of the possible means of mitigation.

1. Introduction

Every time a new system is introduced or an existing system is significantly modified, international standards impose performing a safety assessment to identify if potential new risks are introduced as a result of the innovation. Safety assessments, especially in complex socio-technical domains such as ATM, always require some kind of involvement of people with operational experience (e.g. controllers and pilots) whose knowledge is deemed essential for an adequate understanding and evaluation of risk. However most of the standard safety assessment techniques adopt as a central strategy the use of a safety matrix, aimed at classifying each hazard in terms of its expected severity and acceptable frequency. Despite this method is generally intended to rely on *expert judgment* for an appropriate evaluation of the severity of hazards, operational experts tend to experience difficulties when working at such task. As a matter of fact, the assessment of severity is normally based on so-called *severity classification schemes* which identify a set of *severity classes*. Each class is associated to a different severity level and to a specific outcome (e.g. accident, incident, no safety effect, etc.). As it will be clarified in the following, operational experts cannot easily use these schemes, for at least two main

reasons: (1) Hazards are typically identified as failures to a single function of the system, without considering the potential interactions of such function with other parts of the system and thus without observing it in a realistic operational context; (2) The severity of each hazard is assessed by considering its potential “final” effect, assuming a linear chain of events and infringed barriers which is excessively far from reality.

2. Accident models and risk assessment techniques

In recent years a number of theoretical contributions have investigated the complex nature of accidents in socio-technical and safety critical systems like nuclear power plants, chemical industry and transportation systems. These contributions pointed out the limits of accidents models based on linear sequences of events and cause-consequence configurations. The seminal studies of Charles Perrow (1984) revealed that accidents can be seen as due to unexpected combinations or aggregation of events, named *complex interactions*. More recently Reason’s Swiss Cheese Model (Reason, 1990, 1997) has been considered successful in representing accidents as the result of combined failures at different levels in an organization, including unsafe acts by front-line operators and latent conditions such as weakened barriers (Leveson, 1995, 2004) and defences. Finally, other models like FRAM (Functional Resonance Accident Model) (Hollnagel, 2004) or STAMP (System-Theoretic Accident Model and Processes) (Leveson, 2004) highlighted the emergent nature of failures, which are often the result of dysfunctional interactions between different parts of the system, rather than simple malfunctions of specific components.

Despite these developments in understanding how accidents occur, there has been no comparable development in state of the art risk assessment techniques. Most of these techniques are based on a PRA approach (probabilistic risk assessment), i.e. they adopt as a central concept the well known definition $Risk = Severity \times Frequency$. In such definition both the *severity* and the *frequency* are referred to the potential negative effects of the hazards which can be experienced by a certain system. Thus, in a typical safety assessment, hazards are defined as failures of one or more functions to be mitigated by reducing the frequency of their occurrence and/or the severity of their effects. The overall *level of risk* achieved by the system is the result of the aggregation of the risks identified for each specific hazard.

While this approach is theoretically appropriate for close or simple systems essentially made of hardware components, the application to complex socio-technical systems is problematic, as it relies on a linear representation of hazardous events which is inconsistent with the complex accident models mentioned above.

3. Limits of Probabilistic Risk Assessment

A prerequisite for performing a safety assessment based on a PRA approach is that of identifying a relationship between a set of identified failures for each specific function and a set of possible consequences. This is typically accomplished by filling in FHA tables (Functional Hazards Assessment) and by elaborating them with cause-effects

propagation models, such as FTA (Fault Tree Analysis) and Event Tree Analysis (ETA) (EUROCONTROL, 2003). As anticipated before, such an approach presents a number of drawbacks when applied to complex socio-technical systems. Some of these drawbacks are shortly summarized in the following subsections, by elaborating the contributions of authors like Nancy Leveson, James Reason and Erik Hollnagel.

3.1 Assumed linear link between hazards and their effects

No matter which is the specific graphical notation adopted, PRA typically relies on models of accidents and incidents based on linear chains of events, representing the notion that the preceding event or condition must be present for the subsequent event to occur, i.e. if event X had not occurred then the following event Y would not have occurred (Leveson, 2004). As such these event based models encourage limited notions of linear causality and make difficult or impossible to incorporate non linear relationship (e.g. feedback between system components, including human-human and human-hardware interactions). Furthermore technical and economical constrains impose to focus on the downstream part of the event chain and put emphasis on protection and safety systems that identify a hazardous state after it occurs and then attempt to move the system back into a safe state. As argued by Leveson (2004), this approach -which was appropriate in process industry design (e.g. nuclear power plants)- is largely insufficient in other kind of systems in which emergent configurations of different kind or resources (humans, mechanical, procedural) are essential elements of both the correct and unsafe functioning of an organization (Hollnagel, 2004).

A direct methodological implication is the tendency to assume a linear link between an identified hazard for a specific function (e.g. a technical failure or a human error) and a *final effect* of the hazard itself (e.g. a minor incident, a severe incident or real accident). Such linear link dramatically simplifies the relationship between a failure to a single component and its possible role in producing a negative effect for the safety of the system. I.e. it disregards the well-known notion that a failure to a single function can never be considered as the sole cause of a negative effect for the safety of a system. As we will see in the following section 4, despite this assumption is evident and very strong, the *final effect* is used as a criterion to assess the severity of a specific hazard, considerably influencing the final results of the assessment.

3.2 Initiating events in the chain assumed to be mutually exclusive

A well known limitation of event based models (e.g. Fault Tree Analysis) is that *basic events* are usually assumed to be mutually exclusive. While this assumption simplifies the mathematics in a PRA, it may not match the reality. Leveson explained how seemingly independent failures may have common systemic causes that result in coincident failures (Leveson, 2004). For example, in describing the famous Bophal accident in the chemical industry, she pointed out how a number of components of the chemical plant failed simultaneously (e.g. the vent scrubber, the refrigeration unit, the water spouts and various monitoring instruments). However assigning probabilities to all these apparently unrelated events and assuming independence would lead one to believe that this accident was merely a matter of a once-in-a-lifetime coincidence. A probabilistic risk assessment based on an event chain model most likely would have treated these

conditions as independent failures and then calculated their coincidence as being so remote as to be beyond consideration (Leveson, 2004). On the surface, as she argued, it did seem incredible that these devices were all out of operation simultaneously. Nevertheless, closer looks at the analysis of the accident revealed a quite different picture and showed these were not random failure events but were related to common engineering design and management decisions.

This methodological limitation of PRA is strictly related to the one mentioned before. Assuming the *basic events* as mutually exclusive in the determination of an accident considerably simplifies the task of modelling cause-effect configurations, thus making simpler the numerical definition of the risk associated to each failure at component level. However it can hide critical interactions between different functions or components which are deemed essential for identifying the appropriate mitigation means.

3.3 Risks considered only as functional failures not as dysfunctional interactions

Traditional PRAs focus on *functional failures*, i.e. on the non-performance or inability of specific components to perform their intended functions. However the more complex safety critical systems have become, the more accidents have been determined by so-called *dysfunctional interactions* (Leveson, 2004). Dysfunctional interactions happen when system elements perform as it is expected (i.e. as specified by requirements) but still the overall system behavior results to be unsafe. With the increasing role of human and software in supervisory control, it is quite common to have situations in which a component satisfies its specified requirements, even though the requirements may include behaviour that is undesirable from a larger system context. As studies of organizational accidents in transportation systems have showed (Reason, 1997), accidents happen not only because a pilot or a train driver deviates from a specified procedure or because a component of the aircraft or of the train infrastructure fails to follow the specifications. Actually they can happen because of a critical interaction among different components (electromechanical, digital, human) which was not expected at system design level. In the simpler systems of the past, analysis and testing allowed exercising the system to detect all such potential interactions and changing the system design to eliminate them. In current socio-technical systems a continuous monitoring of emerging critical interactions is required also after the system definition and the system design levels. If the safety assessment is exclusively focussed on functions and component failures, very little insight is produced in order to mitigate the hazardous situations deriving from dysfunctional interactions.

4. The severity classification scheme in SAM methodology

SAM (Safety Assessment Methodology) (EUROCONTROL, 2004) is the standard method for safety assessment in ATM promoted by EUROCONTROL, in compliance with ESARR 4 (EUROCONTROL, 2001). It is made up of three main phases: Functional Hazard Assessment (FHA), Preliminary System Safety (PSSA), System Safety Assessment (SSA) (see central column of Figure 1). The phases are supposed to be performed in parallel with the development and lifecycle of the system under assessment (see left column in Figure 1). This paper is mainly focused on the first phase of the SAM, i.e. the FHA. However some of the considerations made with regards

to FHA also apply to PSSA and SSA. FHA is often performed also during system design and implementation or even during the operational phases, for at least two reasons: a) an existing system is going to be improved, but there is no FHA available as the system was implemented before ESARR 4 was issued, b) practical reasons impose to iterate and refine the FHA, even if the process has reached the stage of performing the PSSA or SSA.

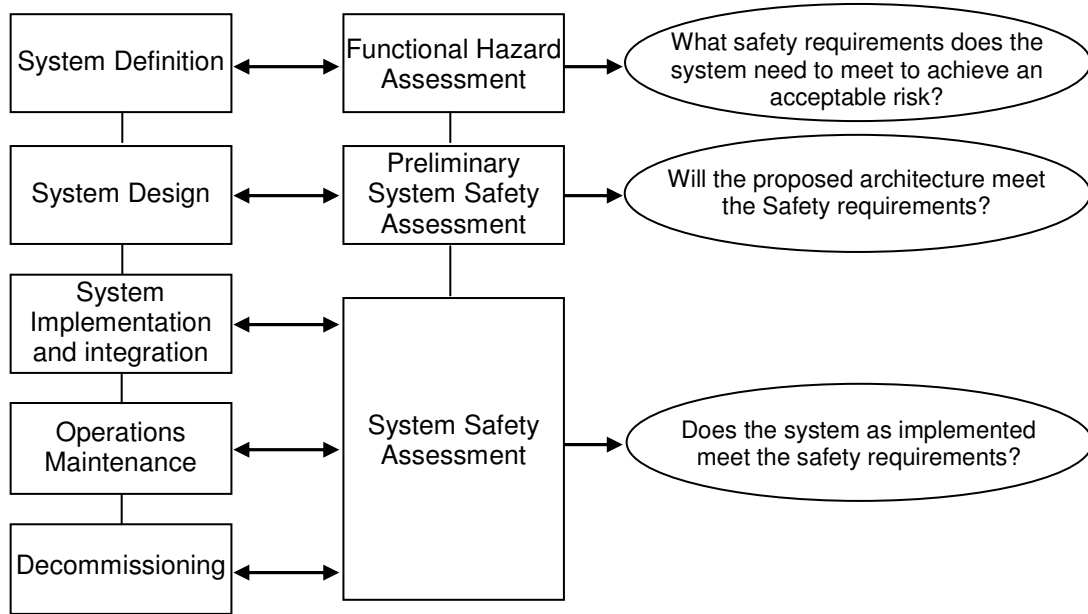


Figure 1: The SAM Methodology

4.1 FHA phases and involvement of operational experts

Main goal of an FHA is specifying a set of safety objectives. These are defined by following five sub-phases:

- 1- Identify all potential hazards associated with the system
- 2- Identify hazard effects on operations, including the effect on aircraft operations
- 3- Assess the severity of each hazard effect
- 4- Specify Safety Objectives, i.e. determine the maximum frequency of hazard's occurrence
- 5- Assess the overall foreseen risk associated to introducing the change or new system.

With regards to the involvement of operational experts, the first three sub-phases are the most important. While for the first step (identify hazards) their contribution is quite easily achieved, the second phase (identify hazard effects) is more difficult and the third phase (assess severity) can become extremely challenging. We will see later in the two case studies how this can even produce nearly arbitrary results, depending on assumptions made. These issues will be clarified in the following, after a brief illustration of the severity classification scheme adopted by the SAM methodology.

4.2 The assessment of severity

As anticipated above, operational experts (typically air traffic controller and/or pilots) are supposed to identify, in collaboration with technical and safety experts, the effect of each hazard identified in phase 1. Effects are then included in textual format in a specific column of the FHA table. Subsequently the experts are required to classify each of these effects in terms of severity, by using the SAM Severity Classification Scheme. The scheme identifies 5 different Severity Classes (SC), from the most severe to the least severe:

- SC1 *Accidents [most severe]*
- SC2 *Serious Incidents*
- SC3 *Major incidents*
- SC4 *Significant incidents*
- SC5 *No Immediate Effect on Safety [least severe]*.

In principle the same hazard can have more than one effect, based on contextual conditions. A typical example is the differentiation of the effects of the same hazard, based on traffic conditions (e.g. low vs high traffic) or weather conditions (e.g. good vs severe weather conditions). However the FHA table should identify a specific SC for each effect, without any particular attention if two effects are produced by the same source hazard. It is to be noted that the SCs are the same adopted in ESARR 2 (EUROCONTROL, 1999, 2000), i.e. they are used by national service providers to classify real occurrences experienced in operational air traffic control centres.

4.3 Problems with the use of severity classes

Experiences made by the authors (see case studies in section 5) suggest that the use of the severity classification scheme brings to the surface some issues in the context of ESARR 4, i.e. in the SAM methodology. While SCs are fit for purpose when reporting and classifying real occurrences as required in ESARR 2, they are very difficult to use when assessing the safety of “pre-operational” systems. Coming back to the issue highlighted in section 3.1, the most problematic aspect is the assumed linear link between a specific hazard and its possible effects. As a matter of fact a specific failure –be it a technical failure or a human error– can never be considered as the sole cause of an accident. I.e. for an accident to occur, a hazard very often combines with several other hazards and contextual conditions, generating a special configuration of elements which cannot be always anticipated when performing the assessment. However, when adopting the functional approach which is typical of PRA, hazards generally does correspond to specific failures. As a consequence, we deal with single failures that could at the same time cause an accident (SC1), different kinds of incidents (SC2, SC3 and SC4) or even no immediate effect on safety (SC5). This is a commonly well recognised point, as demonstrated by the emphasis safety management systems place on near-miss events collection and analysis (Van der Shaaf, Lucas, & Hale, 1991). A near-miss usually shares the same causal factors with real incidents, where mostly contextual (sometimes even fortuitous) factors determine the different outcomes (i.e. no or very limited damage in near miss). Deriving consequences from each single hazard considered in isolation neglects the above reasoning about incident dynamics. In theory, any hazard can result in

a serious incident or even in an accident, depending on contextual conditions and on the way it interacts with other system weaknesses.

Operational experts are normally able to provide very detailed accounts on critical situations for safety and can give valuable insights on the possible consequences of failures or dysfunctional interactions in concrete operational contexts. However, when faced with the task of classifying a single failure in terms of the 5 SCs, they cannot avoid manifesting uncertainties, typically expressed with sentences like “it depends on...it depends how...”. If forced to make a choice, they generally tend to produce classifications that reflect certain assumptions about the contextual situation, or they provide a rationale justifying their answer. Neither assumption nor the rationale will be anyhow considered in the following of the assessment. Another spontaneous strategy is that of ranking the severity with respect to other hazards which have been already classified, despite this choice is theoretically not provided by the methodology. In so doing, experts will generally not take into account the labels identifying the classes (accident, serious incident, etc), but rather reason on a priority ordering, which is independent from the mere labels of the SCs.

It is worth remarking that some attempts to overcome these drawbacks have been actually made in the explanation provided by the SAM guidance material. The ESARR 4 version of the Severity Classification Scheme has been complemented with a set of severity indicators which are supposed to support the assessors when assigning SCs to hazards. These include considerations on the effects of the hazard on air navigation services (e.g. on ACTOs and on flight crew working conditions or on ground ATM and aircraft systems capabilities), on the exposure time, on the number of aircraft exposed and on the actual possibility to recover from the hazard. Associated to each SC, a dedicated text in the Severity Classification Scheme is provided, explaining how these factors should be considered. However, when referred to specific failures and without a reference to a concrete scenario, the guidelines result very difficult to apply and produce the same kind of uncertainties described above.

4.4 Starting from severity or starting from frequency?

The main objective of an FHA is the one specified in the last methodological sub-phase: “Specify Safety Objectives, i.e. determine the maximum frequency of hazard’s occurrence”. This step is accomplished by mean of a Risk Classification Scheme, which is defined by the organization performing the assessment based on ESARR 4 requirements (see an example in Figure 2). The essence of such a matrix is stating explicitly the limits between an unacceptable, a tolerable and an acceptable risk per severity class of an effect.

In theory, this step is only performed after the previous phase – i.e. “Assess the severity of each hazard effect” - has been completed. Based on the severities previously assigned and on the available Risk Classification Scheme, safety objectives are identified in the form of target tolerable frequencies, in order to ensure that the mitigation means identified in the following will adequately mitigated all hazards.

Maximum frequency of hazard effect ("Once per...")		Severity of the effect				
		1	2	3	4	5
Very frequent	< 1 hour					
Frequent	1 hour...5 days					
Occasional	5 days...18 months					
Rare	18 months...150 years					
Very rare	150 years...15.000 years					
Extremely rare	> 15.000 years					

Figure 2: An example of Risk Classification Scheme

Note that the involvement of operational experts is only considered until hazards effects are identified, while no contribution from them is expected when defining safety objectives. In fact the discussion on the severity of hazards should not be affected by the following definition of their acceptable frequency. In practical terms the separation between the two phases is challenged by the mentioned difficulties in the use of SCs by operational experts and by the obvious concern of establishing realistic and sensible targets in the following phase. If an excessive severity is assigned to a certain hazard, there is a risk of over-engineering the system to ensure an appropriate mitigation. On the other hand if a too low severity is assigned to a hazard, there is a risk of forgetting or underestimating it in the following design and development phases.

As we will see, such methodological difficulties are strictly related to the issue of linking each hazard to a specific set of effects and of using the Severity Classification Scheme. The SAM guidance material does acknowledge that the same hazard can have different effects and then different severities. In fact it is specified that SCs should be assigned to the effects of a hazard, rather than to the hazard itself. Besides, different methods are suggested to set the safety objectives, depending on available time and resources. The most complex one (quantitative method) consists in the following steps:

- allocating a different severity class to each hazard effect
- calculating the probability of the hazard to generate each of its effects
- setting the safety objective by choosing the most stringent one, i.e. taking into account not only the severity of the effects, but also the probability of the effect as a consequence of the hazard. The least complex method (qualitative method) consists of identifying - based on expert judgment - the worst credible effect of each hazard in the given environment of operation, then in setting safety objectives taking into account only that effect¹. As a matter of fact, despite the classification of severity should not be influenced

¹ Note that the SAM guidance material proposes 4 different methods for setting safety objectives: Quantitative Method, Prescriptive Method, Criticality Method and Qualitative Method. In the context of

by the following considerations on the acceptable frequency, both the illustrated methods take into account the frequency of hazard effect. Consider also the following quote from SAM methodology, referred to the qualitative method:

The worst credible effect in the given environment of operation should determine the severity class leading to setting of the Safety Objective, using expert judgement. It means that somehow the probability of the hazard leading to certain effect (P_e) has been taken into account when deciding the worst credible severity of the hazard effect.
(EUROCONTROL, 2004, pp. G-14)

According to this quote and to other explanations in SAM guidance material, one could easily conclude that the decision on what is the worst credible effect is strictly linked to considerations on the actual frequency of the hazard. As a matter of fact, there is an implicit recommendation to avoid selecting an effect which is too much severe, unless its occurrence is not considered reasonably frequent. Ignoring such an implicit recommendation is likely to produce too ambitious targets when deciding on safety objectives, which could result not sensible in future steps (e.g. PSSA and SSA) for the organization that is performing the assessment. A practical consequence of these methodological constrains is that economic and design considerations will actually influence the assessment of severity, rather than being a completely independent process.

4.4 Assessing severity and envisaging design solutions

Despite the application of the SAM methodology is expected to be a parallel and independent process from the lifecycle of the system under assessment (see again Figure 1), the analysis made above tends to suggest a different reality. Classifying severity and setting safety objectives without any considerations of economic and technical/procedural constrains does not appear sensible. On the one hand, identifying a safety objective for all possible effects of the identified hazards is clearly too time consuming and expensive. On the other hand, deciding what is the worst credible effect of a hazard influence the definition of safety objectives and - in so doing - constrain future design decisions in the context of PSSA. Furthermore, as we saw in section 4.2, operational experts experience difficulties in using SCs that do not support their way of reasoning on critical scenarios. This might trigger an inconsistent assessment. Practical considerations are likely to induce safety experts to adjust severity classifications and make them consistent with the design solutions which are expected to be adopted. In most cases these are probably checked or discussed with both technical and operational experts to have sufficient insight on the most sensible solutions and to consider what is acceptable in terms of risks. However the contribution of operational experts, rather than being a direct consequence of severity classification, is likely to be more focussed on suggesting priorities and envisaging the consequences of the proposed remedial actions.

5. Two case studies

The case studies presented in this section are both pertaining to experiences in ATM related projects, which include the development of a safety assessment. The first case is the development of an FHA aimed at assessing the improvement of a Short Term Conflict Alert (STCA) in an European military ATC unit. While the second case concerns the overall assessment process of an Airborne Separation Assurance System concept (ASAS), in the context of the European Program “Mediterranean Free Flight”. The reason for illustrating these case studies is twofold. On the one hand they provide evidence of some of the methodological limitations described in previous sections. On the other hand, they document our first attempts to overcome such limitations and propose an alternative approach.

It is worth remarking that our approach is inspired by authors like Erik Hollnagel and Nancy Leveson and by their recent efforts to propose methods more in line with state of the art accident models (e.g. FRAM and STAMP). Hence, we strive to adequately consider in safety assessment the emergent patterns and complex interactions which are typical of socio-technical systems. However, we do not aim at modeling the human behavior with simulation techniques, as some authors have recently proposed (Stroeve et al. 2006). Our main strategy is that of capturing the expert knowledge of the operational personnel (e.g. controllers and pilots), by ensuring their actual involvement at all assessment stages. As most of such knowledge is only available as a *tacit knowledge* (i.e. embedded in working practices), the method we suggest is mainly founded on a scenario based approach (Pasquini & Pozzi, 2005; Pasquini, Pozzi, & McAuley, in press). As we will see in the following, the use of scenarios is essential to achieve a thorough understanding of hazards and of their possible consequences in concrete operational situations.

5.1 Case Study 1: Assessment of a new STCA for a military environment

Case study 1 concerns a safety assessment made in September 2006 for the introduction of an improved Short Term Conflict Alert (STCA) to be installed in a European military ATC unit. STCA is a ground based system intended to assist the controller in maintaining separation between controlled flights, by generating, in a timely manner, an alert of a potential or actual infringement of standard separation minima. The military unit was already equipped with a modern ATC system including an STCA. However, the specific needs of the military environment (military formation flights, aerobatic maneuvers, etc.) created a large number of nuisance alert, rendering STCA ineffective. As the controller working positions made it possible, controllers simply switched the STCA off to prevent the cluttering of information on the radar screen and the well known *cry-wolf syndrome* caused by nuisance alerts.

The safety assessment was part of a wider study promoted by EUROCONTROL, aimed at supporting the military ATM unit in identifying new technical solutions to adapt the STCA concept to the specific needs of the military environment, including a better integration of military and civil airspaces. Examples of these solutions were: a) a *military formation logic*, to minimize nuisance alerts caused by aircraft operating in formation; b) the definition of *region dependent parameters*, to minimize nuisance alerts by taking into account different operational constraints of various areas of airspace; c) specific *lists of*

training SSR codes, to eliminate unwanted alerts caused by training exercises in aerobatic areas, etc. (EUROCONTROL, 2006). As a matter of fact the safety assessment was mainly focused on the possible effects on safety of the new technical solutions, including the possibility of introducing new hazards or unplanned side effects.

Essential part of the safety assessment was an FHA workshop, based on a number of brainstorming sessions attended by 10 people, including safety, technical and operational experts (i.e. military controllers and pilots). Main objectives were:

- Identifying the most relevant hazards
- Understanding their effects on the ATM system
- Assessing the severity of their effects
- Identifying possible mitigation means.

It is worth remarking that in the SAM methodology the identification of mitigation means is expected to be made only at the level of PSSA. On the contrary the FHA should theoretically only encompass the identification of abstract safety objectives pertaining to system functions which are still to be designed. However, in the context of this study, a consideration of possible mitigation means was deemed necessary also at FHA level. On the one hand, no FHA of the previous system was available. On the other hand, the fact that an STCA was already operational (though needing radical improvements) required to immediately reflect on concrete design solutions, in order to mitigate the effects of new possible hazards. Besides, the limited scope and time available for the study did not allow conducting separately an FHA and then a PSSA².

As some of us were involved in facilitating the FHA workshop, a decision was made to integrate it with a scenario based approach, consistently with what already made in the framework of other studies (Pasquini & Pozzi, 2005). As an input to the hazard identification phase, previously to the day of workshop, a set of seven military related scenarios were identified, in collaboration with a controller and a technical expert. In this case the scenarios were textual descriptions of typical operational situations representative of the military environment and airspace under analysis³. The textual description was organized in a tabular format and integrated by a picture taken from a radar screen, representing a specific traffic situation (an example is in Figure 3). Additional cells in the table provided information on the expected behavior of current STCA and on the technical solutions included in the improved STCA, in order to manage the specific situation.

² Due to the limited scope and time available for the study, another simplification with respect to what prescribed by the SAM methodology, is that the FHA has not identified a specific set of safety objectives expressed in numerical terms and has not used a Risk Classification Scheme. As a consequence, the assessment has not included any consideration on the frequency of hazards and on their tolerability.

³ The seven scenarios were named as follows: “Flight in formation/trail”, “Area to Airway”, “Area to area”, “STCA inside and Area”, “VFR/Unknown traffic”, “Authorized penetration”, “Crossing airways”.

OS 2 - AREA TO AIRWAY	
Description	Traffic manoeuvring inside a military area next to a civil airway (ATS routes) with lateral or vertical manoeuvres.
Operational implications	<ul style="list-style-type: none"> ▪ Short reaction time for controllers to react if A/C penetrates civilian airspace. ▪ High speed manoeuvring, high ROC/ROD and steep turns versus steady flight profile. ▪ Aerobatics being performed both by singletons and by formation flights. ▪ Need for ATCOs to input BFL (Block Flight Levels).
STCA implications	<ul style="list-style-type: none"> ▪ Nuisance alerts are generated inside formations. ▪ Nuisance alerts due to excessive prediction times and high speed manoeuvring. ▪ BFL to be taken into account at the CWP ▪ Linear (any) prediction less accurate for the military traffic. ▪ If aerobatics are performed in formation, split tracks can occur.
Technical solution adopted in the new STCA	<ul style="list-style-type: none"> ▪ Creation of buffer zones around aerobatic areas using wider parameters as the Aircraft approaches the boundaries of the area. ▪ Use of BFL as in the current system. ▪ Dynamic activation/de-activation of STCA regions (improved FUA Level3).

Figure 3: Example of a scenario template used during an FHA brainstorming session

The scenarios served at least two different purposes. The first purpose was providing a description of the new system, from an operational point of view. Such description was deemed essential, as most of the controllers were not particularly familiar with STCA functioning in general, due to the mentioned habit of switching off the old inadequate system. The second purpose was supporting the brainstorming for the identification of hazards, by providing a concrete operational context for reflecting on their possible consequences. This purpose, in particular, was an attempt to integrate the functional approach, which requires starting from single functions of the system and thinking about their possible failures. The functional approach was actually maintained, by presenting a functional description of the system and by using checklist with standard prompt words, aimed at thinking about possible failures and human errors (e.g. total loss, partial loss, failure to start, failure to stop, etc.). However the scenarios complemented the functional perspective, as technical failures and human errors were imagined in concrete

situations, allowing engineers and operational experts to derive also more complex hazards, like combination of different hazards or dysfunctional interactions.

First output of the brainstorming sessions was a list of 27 hazards, including a description of possible operational consequences and effects on safety, which were included in a typical FHA table (EUROCONTROL, 2004, Appendix A, pp. A4-A5). Example of hazards were: “Duplicate Mode A”, “Lost Wingman”, “Incorrect military formation detection”, “Incorrect SSR code list input”, “Controller not aware of STCA suppressed for specific aircraft”, etc. In the FHA table, hazards were grouped in a way to keep a reference to the scenario in which they were identified. However, when considered relevant, also the hazards not referring to any specific scenario or identified out of them were included in the table. According to the established method, the hazards identification phase was followed by the assessment of hazard severity and by the discussion about possible mitigation means. At the end all the results were included in the FHA table.

5.2.1 The decision to give up with severity classes

As anticipated in section 4.3, while in the hazard identification phase the workshop attendees were very active in generating ideas and in providing descriptions of the possible consequences of the hazards, much more difficulties were experienced when the experts when confronted with the Severity Classification Scheme. First of all it turned out to be difficult to identify THE specific effect on safety corresponding to each hazard. Then experts stated that none of the hazards would have been the sole cause of an accident, despite nearly all of them could potentially play a role in determining an accident. In addition, the categories *serious incident*, *major accidents*, *significant incidents* or *no immediate effect on safety* were considered difficult or impossible to apply. Even the safety indicators provided in the scheme (e.g. Effects on air navigation services, Exposure and Recovery) were not considered helpful, as the associated descriptions of possible hazards effects are obviously expressed in general and abstract terms: e.g. “partial inability to provide or maintain a safe service”, “significant reduction of functional capabilities” or “hazard may persist for a substantial period of time”. For example defining what is a “substantial period of time” will totally depend on subjective evaluations of the specific operational circumstances experienced and will not necessary imply the risk of producing a serious accident, unless combined with other hazards which cannot be easily predefined.

The limited time available for the workshop (one day and half in total) and the feeling of being stuck with hazard classification resulted in a spontaneous solution directly proposed by some of the attendees. In fact both operational and technical experts had no difficulties in distinguishing between *high severity* and *low severity* hazards, despite they were not able to classify them in terms of the SCs. Particularly they felt the importance of establishing a priority between hazards with an immediate need for a mitigation and hazards that could have been analyzed later, also out of the workshop scope. In other words, they wanted to shortlist a number of candidates for the following methodological phase. As a consequence, despite not rigorous in the terms of the SAM methodology, the proposal of simply distinguishing a set of *high severity* hazards from a set of *low severity* hazards resulted successful. Furthermore, before elaborating proposals

on possible mitigation means (i.e. technical, procedural or training-related solutions) a further distinction was made between hazards the mitigation of which was considered easier and hazards requiring further attention and study. Even though this solution could appear not rigorous in methodological terms, it highlights the strong link perceived between hazards effects on safety and the envisaged design solutions to prevent or manage with them.

5.2 Case Study 2: Assessment of ASAS Spacing concepts in MFF

MFF was a large, six years project, recently concluded, sponsored by the European Union under the TenT Programme. MFF was co-ordinated by ENAV - the Italian Air Traffic Control service provider - and involved several air traffic service providers, especially from the Mediterranean area, and EUROCONTROL. The scope of MFF was to define, test and validate operational concepts and procedures for more efficient use of airspace through the delegation of some tasks related to separation assurance, relying on concepts like Free Routes, ASAS Spacing & Separation, Free Flight (more details can be found below). It focused on the application of those procedures in the particular geographical context of the Mediterranean area, an area located between the core European air traffic area and the States of North Africa and Middle East where a significant growth in demand is expected, especially in the long term. The new operational concepts and related procedures were defined in the early phases of the project (Mediterranean Free Flight, 2001) and their fitness-for-purpose was evaluated through a set of validation exercises, with a iterative process of concept refinement and validation. This included several cycles of Model Based Simulations, three sets of Real Time Simulations (RTS), three Safety Cases, and an extensive set of Flight Trials (FT). Cockpit simulations were used in support to both RTS and FT. More detailed information about MFF are available in (Schäfer & Modin, 2003).

The research issue we faced in this project was mainly due to the experimental nature of procedures and applications to be assessed. The introduction of ASAS procedure deeply changes parts of the existing ATM system, including changes in hazardous conditions and safety issues. Given the novelty of ASAS applications, there was no previous experience on them, nor any existing system with similar characteristics. The safety assessment process was then developed to face two complementary constraints.

(i) Controllers needed to be familiarized with the new procedures and applications, so that they could contribute to the safety assessment as experts. We moved from the assumption that existing operational knowledge still represented a valuable resource, provided that controllers had the opportunity to familiarize with the new system features.

(ii) No experience was available on the system behavior, so a variety of simulation exercises was set up, in order to identify potential hazardous conditions hard to anticipate in the design phase – as it is often the case with those due to unforeseen interactions. These simulation exercises could not replicate the complexity of a real system, but still some system elements could be put in place and observed while working together.

The integration of Real Time Simulations with the safety assessment process seemed a sound solution for both of the above problems (for more information on the MFF safety assessment process and on the use of safety scenarios in RTS, please refer to

Pasquini & Pozzi, 2005; Pasquini et al., in press). The key aspect of such integration was the injection of hazards in the simulation through the implementation of safety scenarios. Only a limited sample of hazards could be simulated (given cost and time constraints), thus hazards were prioritized on the basis of the safety experts confidence in severity and consequences estimates. In other words, the hazard with the highest priority were those for which safety experts were less confident in the estimated severity and consequences. At this point, the major difficulty was to reconstruct a realistic situation where the procedure and the related hazard could be analyzed from a systemic point of view, preserving all contextual factors that shape controller's behavior. Safety scenarios were then used to avoid the assessment of hazards in isolation, so that credible situations could be presented to controllers. The safety scenarios included events such as system failures, pilots and controller errors, and other operational problems (see table 1 below for an example of scenario story board).

Hazard Identification Code: SA2		Airspace Sector: EW
Time		Events
9.48	Accomplice Pilot Action	AZA123 asks to descend to FL290 for technical reasons
9.50	If Accomplice Pilot Action	IBE3674 and AFR432 are cleared to self-separate while crossino
9.50	then Possible Event	AZA123 interfere with IBE3674 and AFR432 (self-separation on-going)

Table 1 - Story board and actions for a safety scenarios

Safety scenarios provide two immediate benefits. First, they give experts an opportunity to reason about what did not work when the system failed and about the potential consequences of the failures. After that familiarization, operational experts are in a better condition to support safety analysts in clarifying some aspects of OHA hazards. This includes defining the realism and the likelihood of some hazards, describing potential consequences, and identifying mitigation means. Further, even if post-simulation debriefings and focus groups are initiated with a discussion on a specific failure/event, operational experts can profit of their level of familiarity with the experimental setting to move to diverse situations and imagine “what if” events, thus widening the scope of hazardous situations considered. The second benefit is that safety analysts have the opportunity to learn through the direct observation of the controller behaviour during the exercises, and to obtain information directly on a series of dedicated events.

However, if we get back to the main line of reasoning of this paper, what was observed during the RTS could not be considered satisfactory as far as the severity assessment was concerned. On a very practical level, we should highlight that of the 5 ESARR 4 levels, the most severe one could never be reached. The highest severity is reached in case of accident, which is simply not simulated in the RTS environment. The closest the simulation can get to an accident is when two aircraft pass one through the other, which in the simulated world results in no damage to any of the two. The two aircraft simply keep flying on their track after “the collision”. More important, the rating

on the other 4 levels could only be based on two criteria, namely (i) percentage of separation infringement and (ii) whether the controller had detected the loss of separation. Both of the two criteria encounter the same drawback we mentioned in the previous section, that is they both address the severity of the end result of an event, which is often the product of highly specific contextual factors. In other words, safety analysts could not simply observe the RTS event and then rate the severity on the basis of the separation infringed and of the controller detection, as this would have implied rating *the causes that had produced the event in the RTS setting* rather than assessing *the severity of a single hazard*. Again, we tried to partially overcome this limitation by profiting of the controllers' expertise. Two workshop sessions were organised after the end of two major simulations, with the objective of reviewing the information gathered on the hazards. Hazards were presented together with the safety scenarios, so that experts could reason about the single hazards not in isolation, but bearing in mind a more realistic situation, that is in interaction with the other system elements. As in the previous case study with the STCA, experts needed to reason about concrete cases in order to draw meaningful estimate on the severity. In the MFF case, scenarios (which had been in a sense validated in the RTS) provided these concrete cases.

The lesson we would like to draw from the MFF case is that the ESARR 4 severity rating encountered difficulties in its application even in a case where it could be applied as a post event classification (i.e. assessing events that were implemented in a simulation). In our opinion, these difficulties stem from the nature itself of the assessment, that is from the fact that experts are asked to assess the severity of an event as representative of a hazard, whilst experts question this very link between hazard and event. They find it hard to trace a linear link between the hazard and the event, and need to draw their estimates from more complex situations, or better said from more realistic situations.

6. An alternative approach to safety assessment

In previous sections we have presented some issues we faced in the safety assessment process, in particular those due to the severity classification scheme. In this section we would like to draw some tentative lessons learnt from the above discussion.

6.1 Assess hazardous situations rather than single hazards

A direct and simple link between a specific hazard and a given effect is a rare case in complex socio-technical systems. In such systems a single failure can normally produce severe consequences for safety only when combined with other factors or failures. It is only the resulting complex configuration that jeopardizes the system defenses. However, it is almost impossible to predefine in formal terms these configurations. They can be somehow anticipated only by means of a thorough operational knowledge. Thus technical failures or human errors are better understood only if analyzed in the context of concrete operational scenarios either describing past events or envisaging future situations. Depending on the aim and scope of the assessment, scenarios can have different formats and uses. E.g. they could be simple narrative descriptions of operational situations to encourage controller and pilots to suggest ideas

during brainstorming sessions, or they could be used as a basis to implement specific exercises during a real time simulation. No matter which is their specific use, they should assure that the hazard identification is performed by technical and safety experts not only in abstract terms.

The traditional functional approach, i.e. consider individually all system functions and imagining their possible failures, is an essential starting point of all safety assessments. Nevertheless it should be always complemented by the analysis of the same events in the context of wider *hazardous situations*, which are better handled and understood by operational experts. Such an integrated approach presents at least two main advantages:

- a) It gives more opportunity to identify not only the more simple functional failures, but also those dysfunctional interactions which generally represent a more insidious threat for the safety of a complex system.
- b) It allows the assessors to work jointly on three different aspects of a traditional safety assessment, i.e. hazards, effects and severity.

Rather than being a methodological approximation, this second aspect is particularly important. In fact the distinction between hazards and effects does not make sense from an operational point of view. What is seen as the causal factor in a certain context can be easily perceived as the consequence in a different one. While, with respect to the assessment of severity, critical scenarios (i.e. hazardous situations) appear as the only meaningful context to express a motivated judgment.

6.2 Prioritize hazards rather than classify severity

A hidden assumption of methods based on PRA and guided by a functional approach is the need and possibility to perform an *exhaustive* assessment of all possible hazards. A corollary of such assumption is that analyzing *all* the single functions of a system and identifying *all* their potential failures will ensure that a complete assessment of risks has been performed. Nevertheless the identification of all potential hazards is far from being a viable solution for a variety of reasons.

First of all, socio-technical systems like air traffic management systems are too complex for a detailed identification of all system functions. Even defining the borders of the overall system - i.e. what is out and what is in the scope of the analysis - is normally a difficult abstraction which cannot ensure completeness. In fact, it usually happens that the scope and the granularity of the system model are defined on the basis of the current aims, often in an almost implicit manner. Secondly hazards do not derive only from failures of single functions but also from dysfunctional interactions among perfectly working functions. These cannot be identified by analyzing each function separately. In addition, due to their emerging nature, they are anyhow difficult to anticipate in pre-operational phases. Last (but not least) the time available for a safety assessment is generally limited in real situation, especially for what concern the involvement of operational experts. E.g. controllers and pilots have generally busy schedules and cannot be attending meetings and brainstorming sessions out of their normal working routine for an excessive amount of time.

Based on these considerations, a detailed classification of each hazard in terms of the 5 SCs appears less important than a careful prioritization of what has been identified. The more hazards are identified, the more productive the assessment will be considered. However, the list of hazards can be never considered exhaustive and there is generally no time available to cover all hazards with a specific safety objective. Thus, it is of paramount importance that the most urgent hazards to be mitigated are identified, no matter which is their rating on the Risk Classification Scheme. In analogy with what has been described in Case Study 1, a subset of hazards can be classified as urgent, to make sure that fundamental design decisions are not taken before these have been adequately considered. Then, also the remaining hazards - at least those which have been considered as relevant - should be recorded, to make sure that they are not forgotten in following design stages.

6.2 Consider safety objectives and mitigation means jointly

A sharp separation between safety assessment and design processes does not appear realistic. From the one end, ensuring that safety is independent from production pressures is an important requisite for the credibility of safety targets. Severe and/or frequent risks should not be hidden or underestimated based on economic or design constraints. In addition, the well known phenomenon of *risk homeostasis* (Wilde, 1994) should be always prevented, in order to ensure that safety improvements are not automatically converted in production benefits, thus reducing the required safety margins. On the other end, looking after safety also means thinking about alternative design solutions, by considering measures on either the technical, the procedural and the training side. Actually, the same safety target can be achieved with different design solutions and with considerable variations in terms of cost and availability. Thus, despite good motivations could justify a theoretical separation between safety and design processes, practical considerations suggest maintaining an adequate communication flow at all stages of safety assessment. This should be assured not only between safety experts and technical experts, but also between the latter and the operational experts. Separation and independence is more a requirement for different organizational functions, rather than a prescribed working method.

The need to consider jointly safety objectives and mitigation means is in contrast with traditional FHA, which is theoretically expected to provide only indications on safety objectives. For example, also in the SAM methodology, mitigation means (or safety requirements) are only considered at the level of PSSA and SSA, as FHA is supposed to reason only in terms of abstract functions, without any speculation on how a specific function will be implemented (e.g. as a hardware component, or as a procedure to be performed by a human operator). Moreover while FHA brainstorming sessions are always required to involve operational experts, the same requirement does not appear explicit in PSSA and SSA.

As for the analysis of hazards and for the identification of severity classes, the approach suggested in this paper goes in a different direction. In our opinion, if the experience of pilots and controllers is deemed essential for identifying the effects of

possible hazards on a system like ATM, it is hardly understandable why their expert knowledge could not be used when considering the possible safety benefits of the envisaged design solutions. This implies a direct consideration of mitigation means also at the FHA level, to make sure that operational experts can actually contribute to the definition of safety objectives. Such contribution should not be limited to a simple choice between abstract categories. Rather, it should influence future design solutions, by considering their effects in concrete operational scenarios.

7. Conclusions

In this his paper we move from the discussion of what appears to us as a fallacy in current state-of-the-art safety assessment, that is the severity assessment seems to blatantly contradict last-generation safety theories. The line of reasoning is then developed by showing the impact of such fallacy in two case studies. We also present some practical solutions we devised to mitigate the issue, but we are well aware that such solutions are mostly *ad hoc* adaptations, far from representing “a solution” to the point we raised.

As a conclusion, we would like to discuss the possible future research direction, in the hope that the above discussion might bear some relevance for our community. In our opinion the key tension we encountered in the safety assessment process is between analytical techniques and a more holistic vision. On one side, we need analytical techniques to pinpoint safety threats. On the other, these analyses “tears the system apart” and tends to overlook the fact that in reality the system elements will work together. To address the actual functioning of the system we then need more holistic techniques, to “reassemble” what we have separated for clarity’s sake. Our proposal is to ground this holistic view in narrative scenarios, to show system interactions as they happen in the everyday functioning. Future research should address the tension between the two polarities – analytical *versus* holistic – and devise solutions to integrate the two perspectives. At the present moment we see the two polarities as representing a contradictory tension we have to deal with, most likely by reflecting on their complementarities rather than opting for one of the two.

Acknowledgements

The authors would like to express gratitude to the EUROCONTROL SPIN Task Force representatives who promoted and supported the FHA study regarding the STCA used in a military environment. Special thanks are due to the ATCC Semmerzake team for hosting the FHA workshop and actively contributing to it.

We would also like to thank all the colleagues of the MFF project for the fruitful collaboration on the activity. The MFF project was partially funded by the EU under the TEN-T program.

8. References

EUROCONTROL (1999) ESARR 2 Guidance to ATM Safety Regulators. Severity Classification Scheme for Safety Occurrences in ATM.

EUROCONTROL (2000) ESARR 2 - EUROCONTROL Safety Regulatory Requirement Reporting and Assessment of Safety Occurrences in ATM.

EUROCONTROL. (2001) ESARR 4 - EUROCONTROL Safety Regulatory Requirement Risk assessment and Mitigation in ATM.

EUROCONTROL. (2003) Review of Techniques to support the EATMP Safety Assessment Methodology.

EUROCONTROL. (2004) Air Navigation System Safety Assessment Methodology (SAM).

EUROCONTROL. (2006). Guidance Material for Short Term Conflict Alert. Appendix D-2: Functional Hazard Assessment of STCA for ATCC Semmerzake.

Hollnagel, E. (2004). Barriers and accident prevention. Aldershot, Hampshire, England ; Burlington, VT: Ashgate.

Leveson, N. G. (1995) Safeware. System safety and computers. Reading, MA: Addison Wesley Publishing Company.

Leveson, N. G. (2004) A New Accident Model for Engineering Safer Systems. Safety Science, 42(4), 237-270.

Mediterranean Free Flight. (2001) MFF Operational Concepts & Requirements.

Pasquini, A., & Pozzi, S. (2005) Evaluation of Air Traffic Management Procedures - Safety Assessment in an Experimental Environment. Reliability Engineering & System Safety, 89(1), 105-117.

Pasquini, A., Pozzi, S., & McAuley, G. (in press) Eliciting Information for Safety Assessment. Safety Science.

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

Reason, J. T. (1990) Human error. Cambridge, UK: Cambridge University Press.

Reason, J. T. (1997) Managing the risks of organizational accidents. Hampshire, UK: Ashgate Publishing Limited.

Schäfer, D., & Modin, E. (2003) A Human Factor Perspective on Free Routing and Airborne Separation Assurance in the Mediterranean Airspace. Paper presented at the 5th International Seminar on ATM R&D, FAA and EUROCONTROL, Budapest, HU.

Van der Shaaf, T. W., Lucas, D. A., & Hale, A. R. (1991) Near miss reporting as a safety tool. Oxford, UK: Butterworth-Heinemann.

Wilde, G. J. S. (1994) Target Risk. Dealing with the Danger of Death, Disease and Damage in Everyday Decisions. Toronto, Canada: PDE Publications.