Sandom C and Fowler D (2006), *People and Systems: Striking a Safe Balance Between Human and Machine*, in Redmill F and Anderson T, *Developments in Risk-based Approaches to Safety*, Proceedings of the 14th Safety-critical Systems Symposium, Bristol, UK, 7 – 9 February 2006, Springer-Verlag.

People and Systems: Striking a Safe Balance between Human and Machine

Carl Sandom,
iSys Integrity,
Gillingham (Dorset), UK

Derek Fowler
Independent Safety Consultant,
Henley on Thames, UK

Abstract

Humans may be viewed as being merely fallible operators of machines; however, that technology-centred view can easily understate the ability of the human to perform tasks which most machines are incapable of doing and to intervene in the event of failure. On the other hand, an overly human-centred view may not take full advantage of the ability of machines to carry out numerically-complex, repetitive tasks consistently and at relatively high speed, and to provide alerts in the event of failure on the part of the human. Somewhere between these extremes lies a more balanced, integrated approach in which the best (and worst) characteristics of human and machine are fully recognised in the development of safe system solutions.

This paper, produced in support of a tutorial entitled: 'System Safety Requirements for People, Procedures and Equipment', given at the Safety-critical Systems Symposium 2006, presents a generic approach for the specification and realisation of safety requirements for both technical and human elements of safety-related systems.

1 Introduction

In the absence of a holistic approach to system safety assessment, it is tempting to concentrate safety assessment effort on what we understand or think we understand (such as hardware and software) and to adopt a 'head in the sand' approach to the human factors which are often perceived as too difficult. Humans are often the major causal factor for hazards in safety-related systems (Sandom 2002) and yet human failures often don't receive proportionate attention in safety analyses. On the other hand, human operators also often provide substantial mitigation between

machine-originated hazards and their associated accidents; yet this too is often overlooked or, conversely, sometimes over-stated.

It is well-established that in some application sectors humans are the major cause of accidents or safety incidents; however, this can lead to erroneous conclusions. Taking the human 'out of the loop' may not be the panacea that it first appears unless we fully understand, for example:

- The potential for equipment failures to cause accidents can be hidden by human mitigation of those failures.
- Humans often perform far less well in monitoring roles than they do if fully involved and occupied.
- Increased automation inevitably leads to de-skilling of the human operator and the ability of the human to mitigate the effects of equipment failure is often impaired.

Apart from a preoccupation with reliability and integrity issues, the development of safety-related equipment is relatively well understood and well covered by process-based safety standards including IEC 61508 (system and software), DO-254 (hardware) and DO-178B (software). However, the role of human factors in system development is far less understood and receives little coverage in the popular safety standards. It is difficult to see how overall system safety can be demonstrated (or even achieved) except through actual operating experience. Safety is not just a matter of system reliability and an argument is made here for safety requirements, including those for human sub-systems, to include functionality and performance as well as the integrity of each safety function.

Some safety-related systems (e.g nuclear reactors) are categorised as such simply because they pose an unacceptable safety risk to their environment and they require additional protection systems to contain that risk within an acceptable level. In contrast, systems such as Air Traffic Control or Railway Network Control are designed specifically to provide risk reduction and can be likened to one big protection system. This paper presents a generic approach for the specification and realisation of safety requirements for the technical and human elements of both types of safety-related systems. The term 'realisation' is used here to cover all activities associated with requirements implementation, validation and verification.

The paper presents a pragmatic methodology to fully integrate human factors analyses with safety engineering analyses to take account of both human and technology capabilities and limitations, thereby addressing the major risks to systems safety. The approach presented here addresses the specification of both Operational-level and System-level safety requirements down to the allocation of functions and safety requirements to subsystems comprising equipment, people and procedures.

However, in order to ensure that such safety requirements are correctly specified, we first need to understand the fundamental nature of safety and safety requirements.

2 Safety Fundamentals

Safety is commonly defined as freedom from unacceptable risk of harm (or accident). One very useful view of safety, and of safety assessment, is the 'barrier model' illustrated in Figure 1 using Air Traffic Management (ATM) as an example.

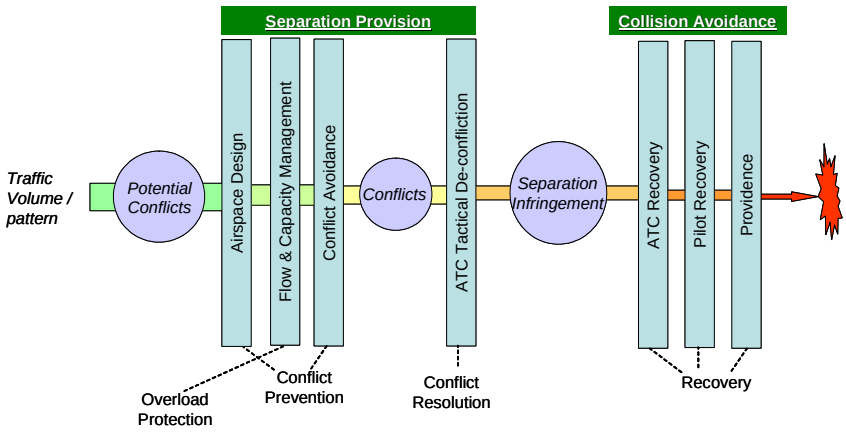

Figure 1. Barrier Model (adapted from Reason 1997)

On the right-hand side of the model is the accident that we are seeking to avoid. In ATM terms, harm is normally taken to be a collision between two aircraft or between one aircraft and a ground-based obstacle – for simplicity we will consider only the case of a possible mid-air collision between two aircraft.

On the left-hand side is the threat posed by the presence of aircraft in the airspace. Intervening between the threat and the accident depends on the presence and effectiveness of a series of barriers. In general, the avoidance of mid-air collisions is dependent primarily on the maintenance of appropriate separation between aircraft or, if that fails, by collision avoidance. Aircraft separation is provided by:

- *Airspace design*: structuring the airspace so as to keep aircraft apart spatially, in the lateral and/or vertical dimensions.
- *Conflict avoidance*: planning the routing and timing of individual flights so that the aircraft, if they followed their planned trajectories, would not pass each other within the prescribed minimum separation.
- *Conflict resolution*: detecting conflicts when they do occur and resolving the situation by changing the heading, altitude or speed of the aircraft appropriately.

In order to prevent overload of the above barriers, the flow of traffic is maintained within the declared capacity of the *Separation Provision* service. *Collision Avoidance* is intended to recover the situation only for those potential accidents that *Separation Provision* has not removed from the system. In general, these may be considered as:

- *Air Traffic Control Recovery* mechanisms – human and/or machine-based safety nets.
- *Pilot Recovery* mechanisms – again, human and/or machine-based safety nets.
- *Providence* – i.e pure chance.

One very important thing that the above barriers have in common is that none of them (neither singly nor in combination) is 100% effective even when working to full specification. This leads us to some crucial conclusions regarding safety, as illustrated in Figure 2:

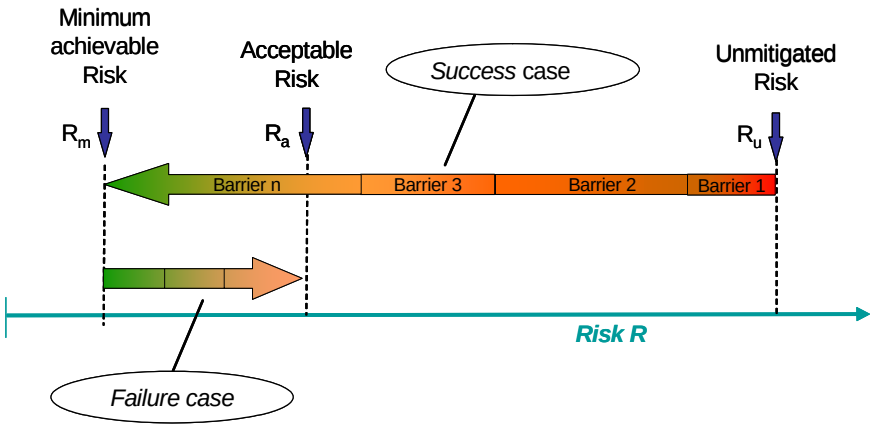

Figure 2. General Risk Model

- Firstly, when risk exists at an unacceptable level (Ru), barriers need to be provided in order to mitigate that risk.
- Secondly, risk cannot be eliminated totally (unless the threat is removed entirely) and the minimum level to which risk can be reduced (Rm) is determined by the desired properties of the barriers – e.g functionality, accuracy, capacity, speed of response etc.
- Thirdly, the risk-reduction effectiveness of a barrier is itself reduced by the undesired properties of the barrier – unreliability, unavailability etc – causing risk to rise somewhat.

Clearly the net risk must lie at or below the acceptable level (Ra). Thus, if we consider a system to include the associated barriers, any safety assessment of that system must address two key issues:

- How safe it is when the barriers are working to specification, in the absence of failure – the *success case*.
- How less safe it is in the event of failure, or partial failure, of a barrier – the *failure case*.

There is a widespread view (unfortunately reinforced by some safety standards) that safety is largely a matter of reliability despite the fact that theory and experience have shown this to be far too narrow a view of safety (see Sandom and Fowler 2003). What the success case tells us is that one of the first considerations in assessing system safety must be whether the functionality and performance properties of the system are adequate to achieve substantially better than an acceptable level of risk.

Once the success case is established, only then is it worthwhile considering the failure case and the increase in risk associated with the failure-related properties of the system. This leads directly to the conclusion that Safety Requirements must take two forms:

- Those relating to the required function and performance, of the barriers – herein referred to as Functional Safety requirements.
- Those relating to the required reliability, availability and integrity, of the barriers – herein referred to as Safety Integrity requirements.

The rest of this paper describes a framework for the specification of Safety Requirements, for a system comprising equipment, people and procedures, using aspects of ATM to illustrate the safety requirements specification process.

3 Safety Requirements Specification

Figure 3 shows a representation of the safety requirements specification process based on a hierarchical framework. An explanation of the five principal levels of Figure 3, appropriate to the development of safety properties, is given as follows:

- The *Operational Environment* (or domain) into which the service is provided. In ATM, the airspace structure and rules, and users of the ATM service, exist at this level and full account must be taken of the properties of the operational domain in the safety specification of the lower levels in the hierarchy.
- The *Service Level*, defined by the barrier model (see Figure 1). Safety targets for the service may be specified at this level.
- The so-called *Abstract Operational Level* at which the barriers that fall within the system boundary are decomposed into abstract safety functions; those safety functions that are entirely independent of whether they are provided by humans and/or equipment. It is at this level that hazards are

defined and *Tolerable Hazard Occurrence Rates* (THORs) are set in order to limit the frequency of occurrence of those hazards sufficiently to satisfy the safety targets.

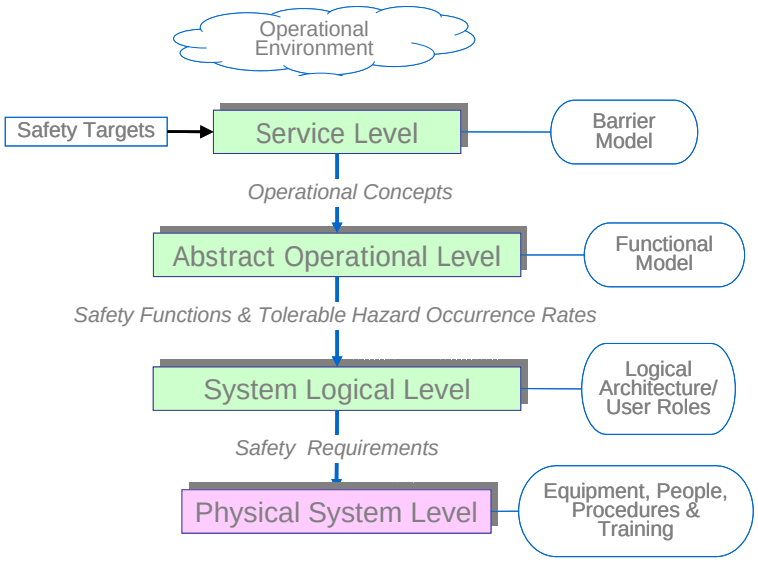

Figure 3. Safety Requirements Hierarchy

- The *System Logical Level* at which the safety functions are allocated to the various elements of the system logical architecture, plus the tasks to be performed by generic human-operator roles – the causes of the hazards are identified at this level, as are the Safety Integrity Requirements that limit the frequency of occurrence of each cause such that the THORs are satisfied; although at this level the distinction between human and machine is made, the safety requirements which emerge from it are still independent of the actual physical implementation.
- The *Physical System Level* - comprising the physical sub-systems, implemented typically in equipment (hardware and software), people (operational and maintenance) and procedures (operational and maintenance). It is at this level that the satisfaction of the safety requirements is demonstrated, possibly via further stages of safety requirements decomposition.

A representation of the relationship between Hazards, Causes and Consequences is the Bow-Tie model, shown in Figure 4, in which all the causes of a hazard are linked directly to the possible outcomes (i.e consequences) in a single structure.

Event Tree Analysis (ETA) is used where appropriate [1] to model all the possible outcomes of a hazard taking account of the mitigations (usually external to the system element in question) that could be used to break an accident sequence should a hazard occur. Working from left to right, each branch of the Event Tree represents a mitigation to which probabilities can be applied in order to express the relative likelihood of success (S) or failure (F) of the mitigation.

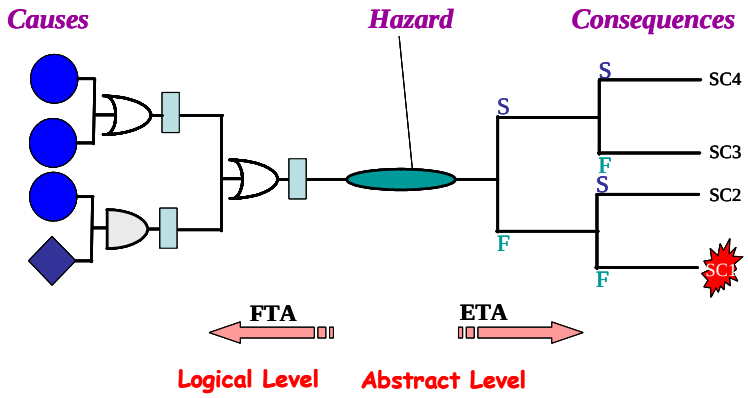

Figure 4. Bow Tie Model

The severities of the various outcomes are categorised - in this case, on a scale of 1 to 4. If safety targets are set for each of these categories, then the THOR for the hazard can be set such that these targets are met, taking account of the probability of success of the various mitigations.

Fault Tree Analysis (FTA) is used to model all the possible ways in which a given hazard could arise from failure within the system element in question, taking account of the mitigations (internal to that system element) that could be used to prevent such failures leading to the occurrence of the hazard. Given the THOR for the hazard, the frequency at which each of the lowest-level events in the Fault Tree are allowed to occur can be determined; each of those frequencies is the Safety Integrity Requirement for that event. The process of developing Safety Requirements is explained in more detail in the following paragraphs.

3.1 Operational Level - Safety Functions and THORs

The first step is to determine what Safety Functions need to be provided at the service level, and to specify the FSR including the performance required of them (e.g accuracy, capacity, timeliness etc, but excluding integrity), in order for safety targets to be met. Figure 5 shows that the Safety Functions are in fact a functional

[1] Usually, ETA is appropriate when there are several possible mitigations for a particular Hazard

description of the elements of the Barrier Model – in this case a simple functional model of the barrier ATC Tactical De-confliction is used to illustrate the point.

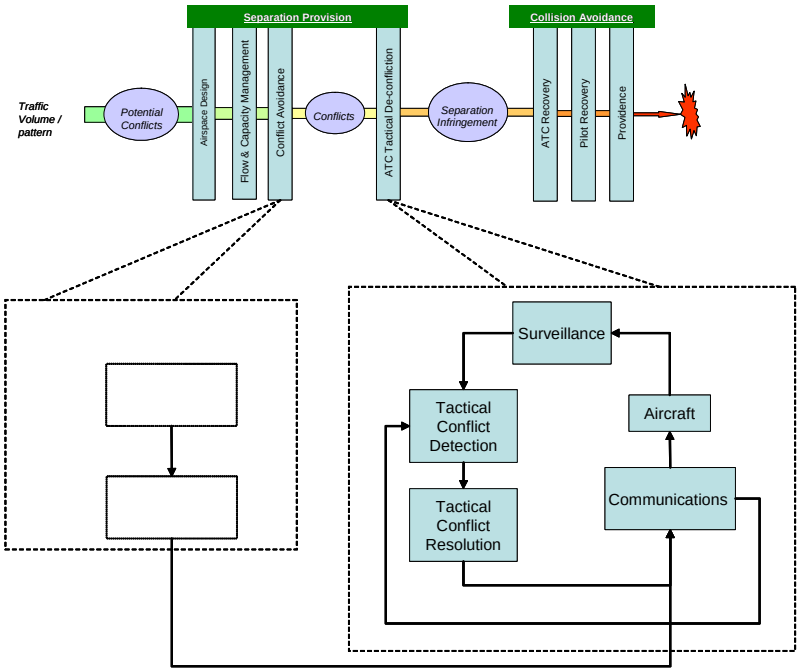

Figure 5. Derivation of Safety Functions

It is necessary at this stage to carry out some form of performance-risk assessment in order to show that specified safety functions are sufficient to reduce the risk to a level (Rm) well below the Safety Targets – i.e minimum acceptable level (Ra) - as indicated in Figure 6. R_a-R_m in Figue 6 represents that portion of the Safety Target, which can be allocated to (functional) failure – clearly these must be a realistic figures otherwise there is no point in proceeding further.

The potential failure modes of the Safety Functions (i.e Hazards) are analysed using the Bow Tie approach, described above, and THORs are specified to limit the allowable rate of occurrence of each Hazard such that the aggregate risk associated with all the Hazards is within the value of R_a-R_m, taking account of any mitigations that are identified during the process.

It is very important in this process that all mitigations are captured as either:

- Additional Safety Functions and corresponding tolerable probability of failure for the provision of deliberate mitigations of the consequences of the identified Hazards.
- Operational Domain Knowledge for any circumstantial mitigations (e.g those arising as a matter of pure chance).

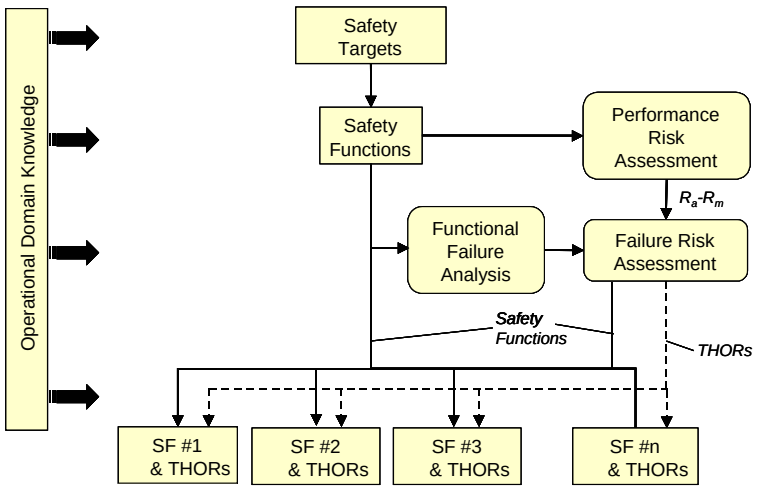

Figure 6. Operational-Level Safety Functions and THORs

3.2 System Logical Level

System Level safety requirements are specified at a logical architecture level – i.e taking into account the distinction between equipment and human elements of the system design but still independent of the actual physical solution.
A generic process for specifying primary and derived safety requirements (the latter through analysing system failure) is illustrated in Figure 7 and it is similar to that for the Operational level, as described above and shown in Figure 6.

Primary system safety requirements stem from an allocation of the service-level safety functions to the subsystem(s) on which they are to be implemented. The example illustration in Figure 7 shows typical ATM equipment sub-systems (Air-Ground-Air communications, Radar Data Processing, Flight Data Processing, and Display) and human-based subsystems (Executive and Planning controllers).

A discussion on the *safe* initial allocation of function between human and machine will be given later in the paper. The hazards and risks associated with failure of each subsystem may be assessed, using the broad Bow Tie approach described above, any mitigations are identified and allocated (as domain knowledge or additional safety functions, as appropriate), and the safety integrity requirements for each subsystem determined.
The safety properties determined from this part of the process being known collectively as derived safety requirements. The outputs from this stage are therefore:

- *Safety functions* to be implemented by each subsystem, and the performance required of them.
- Specification of the *interactions and interfaces* between the subsystems.
- *Safety integrity requirements* for each subsystem.

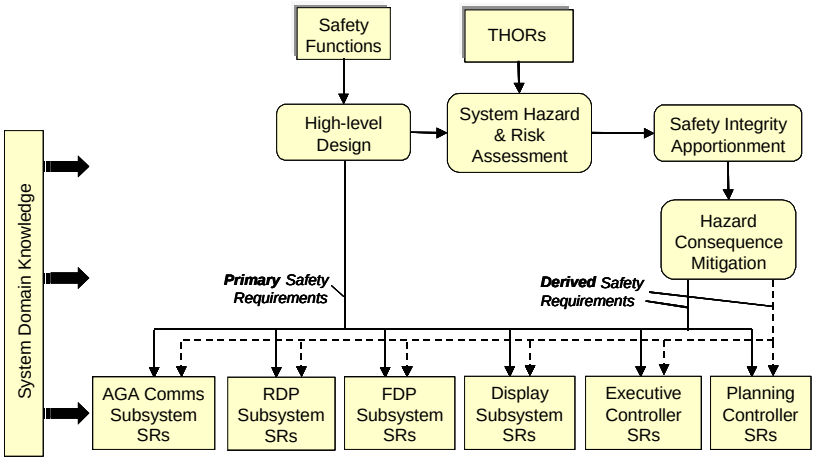

Figure 7. Safety Requirements Specification

A key point here is that the subsystems comprise both technical and human subsystems and the specific methods and techniques used to assess the hazards and risks associated with failure of each subsystem will necessarily be different.

3.3 System Physical Level

As discussed above, the *Physical System Level* comprises the physical sub-systems implemented typically in equipment (hardware and software), people (operational and maintenance) and procedures (operational and maintenance). It is at the physical level that the satisfaction of the safety requirements is demonstrated, possibly via further stages of safety requirements decomposition.

The engineering methods and techniques used for demonstrating the satisfaction of equipment safety requirements (e.g Fault Tree Analysis, Event Tree Analysis, Zonal Hazard Analysis etc.) are relatively well understood by the wider safety engineering community compared with those for people and procedures and will therefore not be discussed further here. The remainder of this paper will discuss how the above approach to safety requirements specification and realisation can be developed in the case of human-based subsystems, using Human Factors methods and techniques.

4 Human Safety Requirements

Human Factors (HF) is a discipline that covers the social, organizational and individual human factors aspects of a system in its context of use (i.e real time). HF analyses primarily address the need to match technology with humans

operating within a specified environment, in order to meet the Operational–level safety requirements.

Previous discussions here on safety requirements have indicated that the scope of system safety analyses must address the system, service and operational environment. This vast scope presents a challenge for the systems engineer who needs to consider the safety-related aspects of the entire system and then to focus the often limited resources available on the most critical system functions.

The human can be both a positive and a negative influence on system safety and humans can alternatively be considered as 'hazard' or 'hero' depending upon the circumstances of the specific system interaction. Ideally, an interdisciplinary approach should be taken to safety-related systems development through the focused application of HF and Systems Engineering methods and techniques – this approach has been referred to as Human Factors Engineering (HFE) (Sandom and Harvey 2004).

4.1 Pragmatic HFE Approach

Broadly, what is required is a *pragmatic* approach to the application of HF methods and techniques for human safety requirements specification at the Logical Level and the demonstration of satisfaction of human safety requirements at the Physical Level. Figure 8 shows different HF analyses that can be undertaken for the specification of human safety requirements (function, performance and integrity) and the realisation of those requirements and their contribution (both success and failure) to safety assurance typically provided by a system safety case.

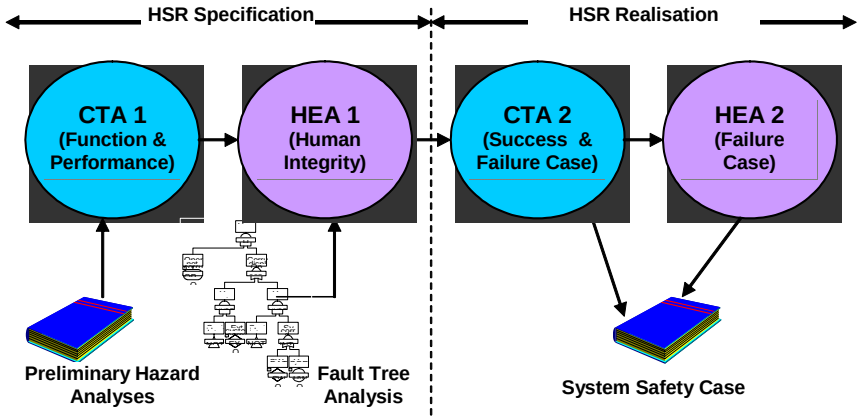

Figure 8. Safety-Related HFE Analyses

Figure 8 shows two different safety-related HF analyses described as Critical Task Analysis (CTA should not be confused here with cognitive task analysis) and Human Error Analysis (HEA).

CTA and HEA are high-level descriptions of analyses which may be undertaken using single or multiple combinations of the various HF Task Analysis,

Human Error Identification or Human Reliability Analysis methods and techniques available.

It is important to note that the CTA deals only with the *safety-critical* tasks and likewise HEA deals only with *safety-critical* human errors. Other HF analyses may have a wider scope to address usability issues which are not directly safety-related. Both CTA and HEA analyses should therefore be planned to ensure that there is no unwanted (and costly) overlap with any wider HF programme.

Typically, two iterations of each analysis should be undertaken to cover human requirements specification and realisation phases and, as the analyses become more focused, the results from each one will inform and focus the other. In addition, these HF activities are entirely complementary as CTA and HEA are bottom-up and top-down analysis techniques respectively (from a hazard to human event perspective).This combination of top-down and bottom-up analyses significantly increases the probability of identifying inconsistencies in the individual techniques and thus enhances safety assurance.

Referring to the safety requirements hierarchy shown in Figure 3, the Operational Level deals with abstract functions with no consideration of implementation details and it follows that there are no specific human factors to consider at that level. CTA and HEA analyses are therefore directed specifically to address the human factors at the system Logical and Physical levels. At the Logical System Level (for each allocated Human SR) safety-related human factors issues may be addressed by undertaking:

- A CTA to validate allocated *human* tasks taking into account procedures and equipment design.
- The specification of human *performance* requirements through an initial CTA.
- The specification of Human Integrity Targets through a HEA of physical system interactions directed by initial system hazard analyses.

At the Physical System Level (for each implemented Human SR) safety-related human factors issues may be addressed by undertaking:

- A detailed CTA to verify human tasks and performance taking into account procedure and equipment design.
- Realisation of HE probability claims through a refined analysis of physical system interactions directed by detailed system Fault Tree Analyses.

4.2 Success and Failure Cases

The Risk Model in Figure 2 makes a clear distinction between success and failure and relates that to the acceptable level of risk at the overall system level using people, procedures and equipment to implement system functionality.

Likewise, a clear distinction needs to be made between the human *success* and *failure* cases as follows:

- The *success* case – the main intention is to assess whether the tasks allocated to the human can be undertaken safely and to identify all the support (e.g procedures, tools etc.) that the human would require while undertaking those tasks.
- The *failure* case – the intention is to identify human error potential and assess reliability when specifically related to the dangerous human errors of commission or omission. In addition, the failure case must identify any human tasks arising from the need to mitigate machine failure.

Figure 9 shows the high-level issues for consideration when making initial decisions relating to the logical safety requirements specification and implementation.

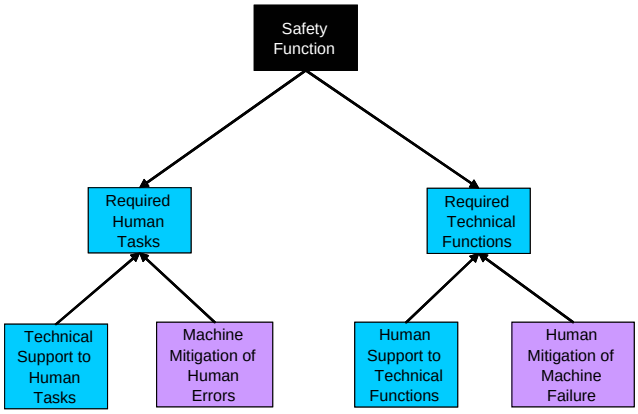

Figure 9. Allocation of Safety Functions

Figure 9 shows the system-level *success case* requirements for tasks and functions which are typically as follows:

- Determination of which Safety Functions should be allocated primarily to the human (as tasks) or machine (as equipment Functions), taking into account the characteristics of the Safety Function.
- Identify what additional human Tasks are needed to support the machine – e.g operation, insertion of data etc.
- Identify what additional equipment Functions are needed to support human performance and achievement of the required Tasks (e.g information, computation etc).

In addition, Figure 9 shows the high-level *failure case* requirement for tasks and functions which can be summarised as follows:

- Technical mitigations of potential human errors.
- Human mitigation.

A summary of success and failure from different perspectives is given in Table 1 and it can be seen that a human success case requires the specification of achievable human tasks to include the successful provision of human mitigation for technical failures where possible.

Case Type	Human View	Technical View	System View
SUCCESS	Human Tasks	Technical Functions	Absence of failure
FAILURE	Human Error (of success tasks AND human tasks for mitigation of technical failures).	Technical Failure (of main functions AND functions for mitigation of human errors	Failure (of Tasks, Functions AND mitigations)

Table 1. Success and Failure Case Summary.

The remainder of this paper will examine the broad issues relating to specific HF methods and techniques that can be used to undertake CTA and HEA which aim to generate detailed evidence to support the human success and human failure cases contributing to the overall system safety assurance.

4.3 The Success Case – Human as Hero

The human success case is built upon the evidence provided by CTA activities undertaken during both the requirements specification and realisation phases of systems development. CTA is a general term applied to the process that identifies and examines task performed by humans, or groups of humans, as they interact with systems. Task Analysis (TA) is a method supported by a number of specific techniques to collect and organize information to build a detailed picture of the system from the human perspective (for comprehensive coverage of TA techniques see Kirwan & Ainsworth 1992). CTA can be used to focus various TA techniques on specific safety issues rather then examining the system as a whole.

CTA seeks to promote appropriate job and task design, suitable physical environments and workspaces, human-machine interfaces and the appropriate selection, training and motivation of the humans involved. At the detailed level CTA examines how the design of human-computer interactions can foster the efficient transmission of information between the human and machine, in a form suitable for the task demands and human physical and cognitive capabilities.

CTA activities can be characterized as being undertaken for one or more of the following broadly defined purposes:

- Allocation of Function.
- Interface design or assessment.
- Task and procedure design or assessment.
- Personnel selection.
- Operability and workload assessment.
- Training requirements or assessment.

For each of these analyses there are specific methods and approaches that are the most appropriate and these are often selected based upon familiarity with the techniques and the aim of the analysis.

The human success case must be built upon two main activities relating to the system safety requirements specification which are the initial Allocation of Function between human and machine and an initial CTA of the functions (or tasks) allocated to the human subsystems to determine what constitutes successful human task performance requirements. Both of these activities are examined here in more detail.

4.3.1 Allocation of (Safety) Functions

The allocation of functions between humans and machines, and defining the extent of operator involvement in the control of the system is a critical activity in safety-related systems. Figure 7 shows a general process for deriving the subsystem safety requirements from a high-level architectural design.

An important feature of Figure 7 is that the high-level design must take into consideration the human factors in the initial allocation of Safety Functions. Too often, this decision is based upon technical capability and the human is allocated whatever functionality can't be implemented in hardware or software, regardless of the suitability of the human to undertake the resultant tasks.

The production of a high-level architectural design requires initial decisions to be made on the allocation of functions to human or equipment sub-systems, in full knowledge of the safety risks involved. Functional allocation decisions need to be informed by good human factors principles and yet the allocation of function is still considered exclusively an ergonomics problem by many systems developers.

The first step is to allocate the abstract operational-level Safety Functions on to the logical model; at this point it is helpful to have a broad notion of how the human and machine will interact in delivering the Safety Functions. The early work of Fitts (1951) was often used to derive MABA-MABA (Men Are Better At-Machines Are Better At) lists that were typically restricted to considerations of either the human or the machine performing each individual function. However, since Fitts' early work, it has become apparent that many functions in complex systems require apportionment of the function between *both* human and machine.

An extensive discussion on functional allocation is beyond the scope of this paper; however, for a detailed review of task allocation techniques see Kirwan and Ainsworth (1992).

4.3.2 Critical Task Analysis

A CTA can be undertaken to identify and analyse the human performance issues in critical operational tasks *as defined for successful interaction*. The initial CTA should focus on human performance aspects relating to the design of the human tasks including high-function cognitive functions such as: attention; vigilance; situation awareness etc.

CTA is a bottom-up technique used broadly to analyse the relationships between system hazards (identified by the System Hazard Assessment in Figure 7) and operational tasks and the HMI design. The analysis works in a bottom-up fashion from operational tasks, related to base events, to identified service-level hazards.

A CTA can concentrate initially on the identification and analysis of the relationships between system hazards and safety-related operational tasks. This analysis will enable both the PHA and TAs to be checked for consistency, providing confidence in subsequent safety assurance claims. Any deficiencies - such as hazards with no related operational tasks or operational tasks (deemed as safety-related by subject matter experts) with no relationship to identified hazards - can be highlighted.

The analysis will also look for opportunities for hazard mitigation through identification of human error potential and improved information presentation by comparing the TA with HMI design guidelines from appropriate sectors. In summary, the CTA will enable the safety-related system developer to:

- Define the allocated safety functions in terms of human operator tasks, including potential mitigations to be provided by the Operator in the event of failure of technical subsystems.
- Capture the interactions and interfaces between the human and equipment subsystems.
- Determine task skills, knowledge and procedure requirements and record these as additional functional safety requirements.
- Confirm feasibility regarding human capabilities performance and reallocate inappropriate tasks to equipment (i.e tools, automation etc) as functional safety requirements.
- Identify training requirements and record these as functional safety requirements.
- Determine human information requirements and human-machine interaction requirements and record these as functional safety requirements.

4.4 The Failure Case – Human as Hazard

The human failure case is built upon the evidence provided by additional CTA and HEA activities undertaken during both the requirements specification and realisation phases of systems development. Broadly, the CTA is undertaken for the specification and realisation of the human tasks (including performance

requirements) required to mitigate against technical failures. The term ‘realisation’ is used to cover all activities associated with requirements implementation, validation and verification. An HEA is undertaken to achieve the following:

- The specification and realisation of Human Integrity Targets relating to the success-case human tasks.
- The specification and realisation of Human Integrity Targets relating to the human tasks required to mitigate against technical failures.

Human subsystems must be specified and acceptable Human Integrity Targets specified for the identified sources of human error. In addition, the validation and verification of the achievement of the allocated Human Integrity Targets for each human subsystems is also required (this may include procedures as well as people).

Figure 7 shows a generic process for deriving both technical and human system-level safety requirements from a high-level architectural design. However, the specific processes for determining primary safety requirements and producing derived safety requirements will necessarily be based upon on different analysis techniques when dealing with human rather than technical subsystems.

Figure 10 (an adaptation of the generic Figure 7) shows the high-level process for deriving system-level safety requirements for humans using HEA to generate integrity requirements based upon an analysis of human failure.

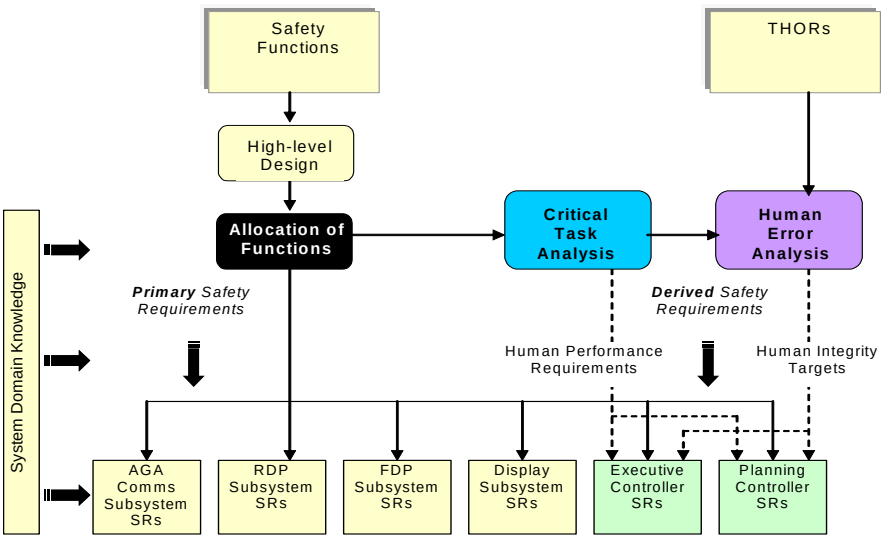

Figure 10. Human Safety Requirements Specification

HEA analysis is a top-down technique used to model the relationship between critical human failures and hazards, and the mitigating aspects of the system design.

An HEA should be undertaken using a two stage process of Human Error Identification (informed by the CTA) followed by a Human Reliability Assessment (informed by other safety analyses such as FTA etc.) which can be either qualitative or quantitative as required. Both of these activities are examined here in more detail.

4.4.1 Human Error Identification

Historically, the emphasis in Human Reliability Analysis (HRA) has been on techniques for the derivation of Human Error Probabilities (HEPs) for use in systems analysis techniques such as FTA. However, HEA should be an integrated process that includes a systematic and rigorous qualitative analysis to *identify* the nature of the errors that can arise prior to any attempt at quantification. This qualitative Human Error Identification (HEI) must ensure that no significant failures are omitted from the analysis.

It is widely recognised that there are considerable uncertainties in the quantitative data available for inclusion in HRA. However, as long as the qualitative error identification process is sufficiently comprehensive, valuable insights will emerge with regard to the sources of risk, and where limited resources should be most cost effectively applied in minimising these risks.

4.4.2 Human Reliability Analysis

The derivation of quantitative human integrity targets is difficult and HRA techniques have attempted to address this issue (see Kirwan 1994). However, much of the HRA research has been dominated by assumptions that apply to technical systems and arguably these do not translate well to human systems. While the failure probability of hardware can be largely predicted by its basic design and its level of use, human error probabilities are influenced by a much wider range of contextual factors, such as the quality of the training, the design of the equipment and the level of distractions.

The terms ‘Performance Shaping Factors’, ‘Performance Influencing Factors’ or ‘Error Producing Conditions’ are often used interchangeably to refer to the direct and indirect factors that influence the likelihood that a task will be performed successfully.

A pragmatic method of addressing this issue is to undertake a HRA focused specifically on the basic human events identified by the system safety analyses and in particular from the system Fault Tree Analyses. For systems, which typically have a high degree of operator interaction, many basic FTA events will be identified as human interactions. Once each fault tree is modelled, predictive, quantitative failure data can be input at the bottom from Availability and Reliability data for all hardware and software base events. By subtracting these

values from the associated hazard target, quantitative Human Integrity Targets (HITs) can then be calculated for each critical human event.

An HEA would then focus on developing specific safety arguments for each basic human event to provide evidence that the HITs can be achieved.
For critical areas, where the HEA reveals that the HITs are unrealistic, mitigations can be re-assessed and recommendations developed for further action. In this way, no predictions are being made about the human error rates; rather, the HITs are derived from the remaining integrity requirements once the hardware and software failure data is input and an analysis is undertaken to ascertain if the remaining human integrity requirements are realistic.

5 Conclusions

This paper has examined problems associated with the specification and realisation of functional safety requirements for the human elements of a system for which a target level of safety is specified at the service level. It was shown that the high-level allocation of functions to hardware, software or humans must be done by taking human performance and limitations into account and a generic approach was presented for the specification of both service-level and system-level safety requirements down to the allocation of functions and safety requirements to subsystems.

The process for the specification of human subsystem safety requirements is no different to software or hardware; although it is arguably considerably harder due to the difficulties associated with the immense scope and variety of issues affecting the reliable performance of human tasks. This paper has examined issues relating to the consideration of human subsystem safety and has outlined the scope and activities necessary for a comprehensive human factors safety analysis. A pragmatic method was introduced that advocates the application of focused Human Factors techniques to the assurance of safety for human subsystems.

The relative difficulties associated with the specification, implementation, validation and verification of human safety requirements, compared with safety requirements for hardware and software, should not be underestimated and this paper has not addressed many of these difficulties in detail. However, this paper has outlined a high-level approach for a focused and integrated application of Human Factors analyses for the specification and realisation of human subsystem safety requirements.

References

Fitts P M (1951). Human Engineering for an Effective Air Navigation and Traffic Control System, National Research Council, Washington D.C., 1951

Kirwan B (1994). A Guide to Practical Human Reliability Assessment, Taylor & Francis, 1994

Kirwan B and Ainsworth L K (Eds.) (1992). A Guide to Task Analysis, Taylor and Francis, 1992

Reason J (1997). Managing the Risks of Organizational Accidents, Ashgate Publishing

Sandom C and Harvey R S (Eds.) (2004). Human Factors for Engineers, IEE Publishing, London

Sandom C and Fowler D (2003). Hitting the Target - Realising Safety in Human Subsystems, Proceedings of the 21st International System Safety Conference, Ottawa, Canada, August 2003

Sandom C (2002). Human Factors Considerations for System Safety, in Components of System Safety, Redmill F and Anderson T (Eds.), proceedings of 10th Safety Critical Systems Symposium, 5th-7th February 2002 Southampton, Springer-Verlag, UK, February 2002