IMPROVING RAILWAY SAFETY BY PROPERLY MODELING "COMPLEX SOCIO-TECHNICAL SYSTEMS”

Jorge Almeida, Alexandra Fonseca

Principal Engineer, Senior Engineer

Critical Software S.A.

SUMMARY

It is commonly understood that critical systems impacting railway safety are subject to a rigorous verification and validation (V&V) process, requiring certification before being accepted for commercial use. So, why do accidents still seem to occur more frequently than expected?

Accidents like the one near Santiago de Compostela, Spain, in July 2013 make us think how far human factors are being sufficiently and properly covered by railway standards in system development and commissioning. A set of relevant questions were raised at the time: "Why was the driver going at that speed? Was there a train control system in operation? If so, why was it not working? Otherwise, why was there not a control system in place?”.

This paper discusses the actual role of human factors in Reliability, Availability, Maintainability and Safety (RAMS) analysis according to the CENELEC standards specifications. This paper also presents suggestions on how these specifications need to evolve in order to properly assess “complex socio-technical systems”, focusing on identifying human errors and creating appropriate safety barriers in the system design to reduce the impact of human factors in the final evaluation of risk.

INTRODUCTION

Railway accidents are often the result of a combination of factors, frequently involving human error in some shape or form. Even when using state-of-the-art signalling and protection systems, accidents still happen mainly because human behaviour is unpredictable. A supposedly highly-dependable train safety control system, where the human interaction elements are not properly assessed, can easily be compromised and lead to unsafe situations.

Although CENELEC railway safety standards (applied worldwide) already specify that “the achievement of railway RAMS requires more rigorous control of human factors, throughout the entire system lifecycle, than is required in many other industrial applications”, the facts show that there are still gaps regarding this subject. These systems must be assessed taking into consideration several aspects that are not the main focus of traditional RAMS methodologies.

This paper presents some approaches already applied in other industries to define appropriate human error reduction strategies and how they can contribute to improve safety in the railway industry. After outlining the challenge, the paper assesses the human impact in railway accidents, how the appropriate standards are dealing with these factors and the existing techniques to perform Human Reliability Analysis (HRA). Finally, we propose the application of the STAMP methodology, already in use in other industries, to the railway domain, in order to effectively assess the dynamic nature of actual complex socio-technical systems.

UNDERSTANDING THE CHALLENGE

Even though human errors are identified as one of the main causes of railway accidents [1], there is a significant lack of accurate estimations for the values to be used in a quantitative risk analysis taking into account human factors. The error probability of every human action is often rated with the fixed value 10^{-3} [14], but there are other values obtained from more complex risk analyses that present different values ("Rasmussen's three levels of behaviour: knowledge-based, rule-based and skill-based", 1983 and "Values for human error in railway transport" published by Hinzen, 1993). However, none of them are fully proven to be valid for use in RAMS calculation for the railway transport, because human errors have a holistic nature and any risk analysis must follow an approach taking this fact into account, and cannot be seen only from a quantitative perspective.

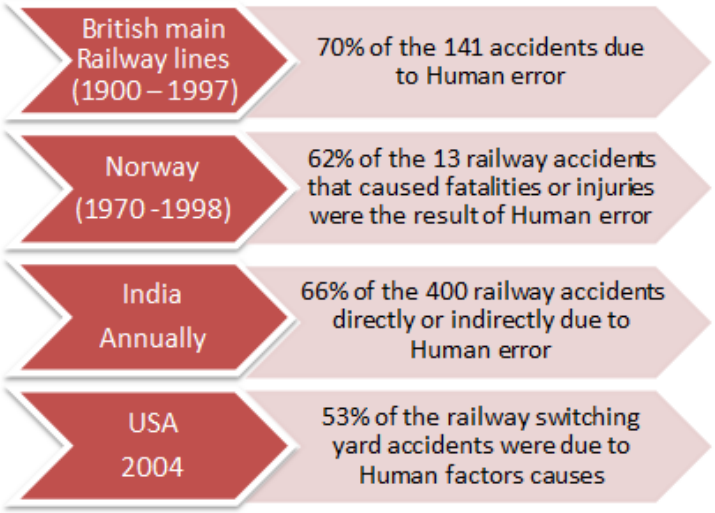

Figure 1: Examples of human error impacting railway accidents [1]

THE ROLE OF HUMAN ERROR IN RAILWAY ACCIDENTS

Some studies have already been undertaken in order to evaluate the human factors contributing to rail accidents. This section presents two of those analyses, one of them based on the railway domain, the other on aviation.

Baysari analysis

Through an analysis to understand the human factors contributing to railway accidents and incidents in Australia [3], Baysari describes the errors associated with the "human failure" incidents analysed from the period 1998-2006, taking into account 40 investigation reports. The conclusion were as follows:

- *Unsafe acts* – The most common error types were skills-based errors. Of these skills-based errors, most were the result of an attention failure;
- *Preconditions for unsafe acts* - The most common problem was the perception of an incorrect expectation/assumption;
- *Unsafe supervision* - The most frequent problem was found to be inadequate supervision, specifically a failure of supervisors to track worker performance;
- *Organisational influences* - The most common problem was inadequate equipment design.

The global analysis of "human failure" investigation reports exposed that **skills-based errors were the most common errors.**

Wiegmann and Shappell analysis

The *Human Factors Analysis and Classification System* (HFACS) from Wiegmann and Shappell, 2003, describes four levels of failure (Figure 2) based on the "Swiss Cheese" model of human error and defines categories inside each level. HFACS distinguishes between the "active failures" of unsafe acts, and "latent failures" of preconditions for unsafe acts, unsafe supervision, and organizational influences. These categories were developed empirically on the basis of over 300 aviation accident reports [16].

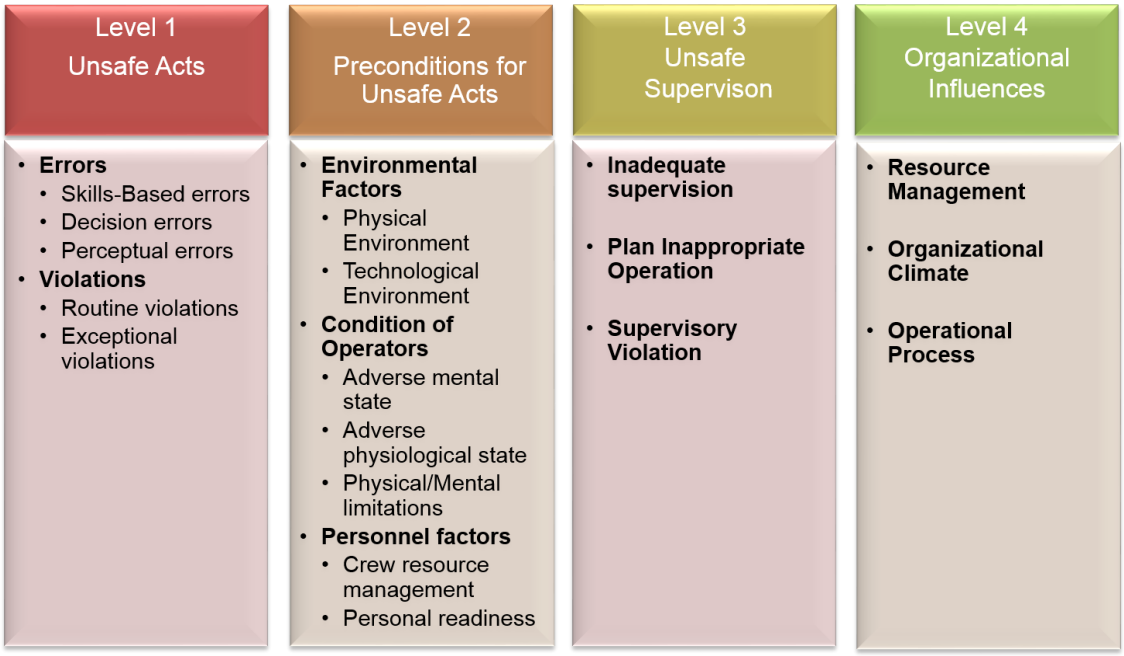

Figure 2: HFACS (Wiegmann and Shappell, 2003)

Both analyses present similar conclusions regarding the most common types of human failures and the conditions or influences that lead to those failures. These categories can be considered as a baseline for a human factor assessment taxonomy.

HOW RAILWAY STANDARDS DEAL WITH HUMAN FACTOR

CENELEC European railway standards are being used not only in Europe but worldwide. These standards, particularly EN50126/EN50129, clearly specify that human factors must be considered throughout the system life-cycle. EN 50126 identifies some human factors that should be addressed because they can influence the system development process. Following the management process for RAMS described in the EN50126 standard, EN50129 mandates the effective implementation of:

- *Quality management* - in order to minimise frequency of human errors at each stage in the life-cycle, and thus to reduce the risk of systematic faults in the system, sub-system or equipment.
- *Safety management* - in order to reduce the instances of safety-related human errors throughout the life-cycle and thus minimise the residual risk of safety-related systematic faults.

EN50129 also specifies that one of the key aspects for a given system achieving its required safety functions is the systematic failure integrity which is not quantifiable and is related to hazardous systematic faults caused by human errors in the various stages of the system/sub-system/equipment life-cycle. According to the standard, the

required systematic failure integrity is achieved by means of the quality management and safety management conditions effective implementation and by applying the specified technical defences against these types of fault. Since it is not possible to use quantitative methods to assess systematic failure integrity, the standard proposes Safety Integrity Levels (SIL) to group methods, tools and techniques and considers that if they are effectively applied it is possible to provide an appropriate level of confidence in the realisation of a system to a stated integrity level.

However, both EN50126 and EN50129 do not “clearly” define some concepts related to human factors influencing different system life-cycle stages. EN50126 defines the five human factors in Figure 3 to be assessed in system design and development. EN50129 presents some techniques, measures and guidelines to manage those factors.

EN50126 Human Factors	EN50129 Techniques / Measures / Guidelines
Human Competency	• Qualification of staff in safety organization • Training of staff in safety organization • Lack of domain knowledge
Human independence during design	• Proposes arrangements for independence between teams/roles depending on the SIL • Unclear competence requirements for each independent role
Human involvement in Verification and Validation	• Audit of processes by safety organization • Establishes indepence between designer of test facilities and designer of the system or product • Lack of domain knowledge
Interface between human and automated tools	• Tools must be proven in use or validated • Unclear definition of interface Human/Tools
Systematic failure prevention processes	• Audits • Intensive and rigorous testing activities • Lack of domain knowledge

Figure 3: Approaches to human factors in EN50126 and EN50129

One can highlight from Figure 3 the lack of consideration for domain knowledge in the assessment of the defined human factors. There is clear evidence, from studies performed in other domains, that the most significant factor in reducing hazardous failure rates is better domain knowledge [4].

Furthermore, it is important to realize that human behaviour can vary despite being presented with similar situations and that human understanding and interpretation about a concept that is not well specified will vary substantially depending on factors such as the environment, human motivation and culture.

So the standards must be more effective in defining the degree of **human competence** required for each stage in the system life-cycle and competence requirements for each role, whilst specifying the need for **domain knowledge**. Additionally the **interface human/tools** should be specified.

HUMAN RELIABILITY ANALYSIS (HRA)

The reliability of a system depends not only on the reliability of software and hardware components but also on human reliability, which is usually defined as the probability that a person will correctly perform some system-required activity during a given time period (if time is a limiting factor) without performing any extraneous activity that can degrade the dependability of the system [2].

In order to overcome the lack of a proven value to quantify the human factor in risk analysis, some techniques have been developed to perform Human Reliability Analysis (HRA). The purpose of the HRA is to estimate the probabilities of human errors that can potentially have an impact on safety. It aims to calculate the nominal Human Error Probability (HEP) by using Performance Shaping Factors (PSF).

Existing techniques can be split essentially into two categories (Figure 4 and Figure 5), first generation techniques and second generation techniques. The latter are more theory-based regarding assessment and quantification of errors; they are based on a cognitive model which is more geared up to explaining human behavior. To give a better understanding of the difference, it is important to have a definition of human cognition: "*the act or process of knowing, including both awareness and judgement by an operator*" [5].

First generation techniques	Advantages	Disadvantages
• **APJ** - Absolute Probability Judgement • **HEART** - Human Error Assessment and Reduction Technique • **JHEDI** - Justified Human Error Data Information • **PHRA** - Probabilistic Human Reliability Analysis • **OATS** - Operator Action Tree System • **THERP** - Technique for Human Error Rate Prediction ↓ **Most popular and effectively used**	• Easy to use • Highly quantitative aspects • Highly flexible and applicable in a wide-range of areas	• Ignore the cognitive processes that underly human performance • Do not consider the impact of factors such as environment, organisational factors, and other relevant Performance Shaping Factors (PSF) • Do not consider errors of commission (incorrect performance of an assigned action.) • Focus on omission errors (actions not performed) • Binary representation of human actions (Success/Failure)

Figure 4: HRA – First-generation techniques advantages/disadvantages [2] [15]

Second generation techniques	Advantages	Disadvantages
• **CES** – Cognitive Environmental Simulation • **CAHR** – Connectionism Assessment of Human Reliability • **MERMOS** – Méthode d'Evaluation de la Réalisation des Missions Opérateur pour la Sûreté • **ATHEANA** – A Technique for Human Error Analysis • **CREAM** – Cognitive Reliability Error Analysis Method ↓ **Most popular and effectively used**	• Allows to follow the temporal and logical process that leads to manifestations of inadequate behavior. • Performance Shaping Factors (PSF) were derived by focusing on the cognitive impacts on operators. • Flexible to be applied in different domains.	• Not able to enclose a complete accident sequence in which different situations of error/malfunction occur and establish a trajectory for an accident. • Need better human behavior modeling.

Figure 5: HRA – Second-generation techniques advantages/disadvantages [2] [15]

From the above tables one can highlight the gap that still exists in **connecting human performance and human culture** in evaluating how each individual will realize that he/she is facing a hazardous situation and how he/she will react to that situation, because two different people will behave differently in a similar situation. This leads to the need for **better human behaviour modeling**.

THE WAY TO FOLLOW

The above sections presented the gaps in the CENELEC standards and in the HRA techniques regarding human factor evaluation, identifying the main factors that influence an effective risk analysis about human contributions to rail accidents. The following main issues were highlighted:

- Human competence
- Domain knowledge
- Human - tool interface
- Human performance / human culture
- Better human behaviour modeling

Furthermore, the increasing complexity of the rail networks makes it more imperative to ensure full coverage of all risk situations and guarantee that human factors are reliably managed in risk assessment activities.

So, new strategies should be adopted in order to manage human errors and guarantee that human factors are properly assessed by ensuring that hazardous situations are detected long before they can emerge in an accident situation.

COMPLEX SOCIAL-TECHNICAL SYSTEMS

Definition

The operation of “complex social-technical systems” recognize that one cannot disassociate human interactions and decisions from the influence of the physical system itself or from the organizational culture in which the human operator is working.

Modern complex systems are composed of several sub-systems which interact and collaborate with each other. These interactions (emergent behaviors) may generate accidents that are not caused by any component failure, and therefore are hard to model with traditional RAMS methodologies.

Identifying all of the errors that can lead to an accident is not a straightforward task because the same events can lead to widely different consequences.

In order to properly evaluate all of the dimensions of these types of systems with accidents generated by non-linear interactions, alternative approaches to accident modeling have been proposed focusing on building a strong safety control program rather than focusing on finding the root causes of the accidents. These models are already being applied in other domains successfully, so their adaptation to the railway industry should be possible.

Traditional modeling approaches

Several accident modeling methodologies have been proposed to perform risk assessment activities for safety critical systems. Some of them concentrate on component failure which makes it difficult to model all the interactions and dynamics of modern complex socio-technical systems, such as:

- *Sequential Accident Models* - Present the accident as a sequence of discrete events caused by a specific failure (human or equipment) that occurs in a particular order. The main objective is to improve the reliability of the least reliable system components (e.g. FTA, FMECA).
- *Epidemiological Accident Models* - Presents the accident as a consequence of missing or degraded barriers. The main objective is to identify and implement new barriers or improve the existing ones (e.g. SOAM).

In order to fill the gaps of these approaches in human factor assessments, another approach emerged, the Cognitive Systems Engineering Approach, based on the modeling of human-machine systems. It considers that, in order to understand what can go wrong, one must understand the way joint human-machine systems work nominally and determine the variability of the interaction of such systems (e.g. ATHEANA, CREAM). However, as previously described, there is still a gap in these approaches when it comes to connecting human performance and human culture.

To cover all the dimensions of complex socio-technical systems that traditional models are unable to, another approach has been proposed, the *Systems Theoretic Approach,* wherein interaction between systems are modelled as feedback loops of information and control. Accidents are caused by instability in the feedback loops. One of the most important aspects in this approach is that the system is not considered static but rather a dynamic process that shall accommodate changes in the environment and itself. One of the most notorious methodologies following this approach is STAMP (Systems-Theoretic Accident Model and Processes) [6].

STAMP (Systems-Theoretic Accident Model and Processes)

The STAMP methodology has been proposed by Nancy Leveson (2004) [7] to tackle the accident modeling of complex socio-technical systems. This model considers the technical (hardware and software), human and organisational aspects of systems.

STAMP approaches safety as a control problem where accidents are seen as a result of interactions between system components that break system constraints. Since STAMP is all about control processes, it naturally handles the dynamic nature of complex socio-technical systems. Leveson (2008) states that STAMP considers "*social and organisational factors, such as structural deficiencies in the organisation, flaws in the safety culture and inadequate management decision making and control*". "*Human error is treated as part of an on-going process that is influenced by context, goals, motives and mental models*". STAMP changes the emphasis in system safety from preventing failures to enforcing behavioural safety constraints [7] [8].

Why should the STAMP model be considered in RAMS assessment?

- Accidents involve complex dynamic processes. They are more than a linear sequence of events.
- STAMP enforces constraints on component behaviour and interactions to prevent accidents.
- It captures more causes of accidents by considering unsafe interactions among components, design errors, flawed requirements, complex software behaviour and **complex human behaviour**.
- Assesses accidents as being a control problem and not a failure problem, meaning that **accidents are caused by inadequate control**.
- Most of the accident models are focused towards the operations phase. This means that they were designed to analyse accidents that have already occurred. STAMP can be tailored in order to be applied early in the system life-cycle and be used as support of the development phase.
- STAMP is already supported by commercial tools, such as “SpecTRM” from Safeware Engineering Corporation™. This is a major advantage for safety and reliability methodologies since it allows a more efficient execution of often effort-intensive activities.

How to identify inadequate control in a system?

- Through the STPA (System-Theoretic Process Analysis), a new hazard analysis technique built on STAMP, which focuses on addressing the development phase of a safety-critical system, and identifies potential hazards that might lead to accidents. STPA identifies the same hazardous scenarios as fault trees, but also additional ones involving complex software and human errors.

How to identify inadequate control that can cause an accident?

- Through CAST (Causal Analysis based on STAMP) , a framework built to assist in understanding the entire accident process, identifying the most important systemic causal factors involved [8].

For this paper we will focus only on the STPA technique and in how it can contribute to support the system development phase by identifying the kinds of inadequate control that can lead to an accident early in the system life-cycle, creating safety constraints to eliminate or control unsafe control actions.

STPA (System-Theoretic Process Analysis)

STPA is a new hazard analysis technique built on STAMP, with the aim of identifying system and component requirements and constraints to be used in design phase. STAMP begins by considering the four types of hazardous control actions that must be eliminated or controlled to prevent accidents:

1. Control action not given
2. Control action given incorrectly
3. Control action given at the wrong time (too soon, too late or out of sequence)
4. Control action stopped too soon or applied for too long

These four types of hazards are similar to the ones used in FMEA (Failure Mode and Effects Analysis) technique. Based on these four types of hazardous control actions, STPA follows the methodology presented in Figure 6.

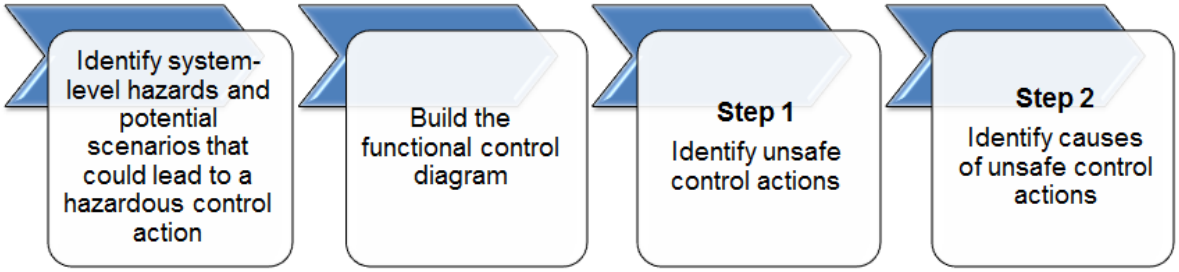

Figure 6: STPA Hazard Analysis

The generic high-level functional control diagram that supports the STPA is presented in Figure 7. The red boxes present the scope of Step 1 and Step 2 from Figure 6.

The controller's process model aims to identify the environmental and system states that can affect the safety of the control actions. This way the controller will be able to make safer decisions, otherwise it cannot be designed to provide safe control actions.

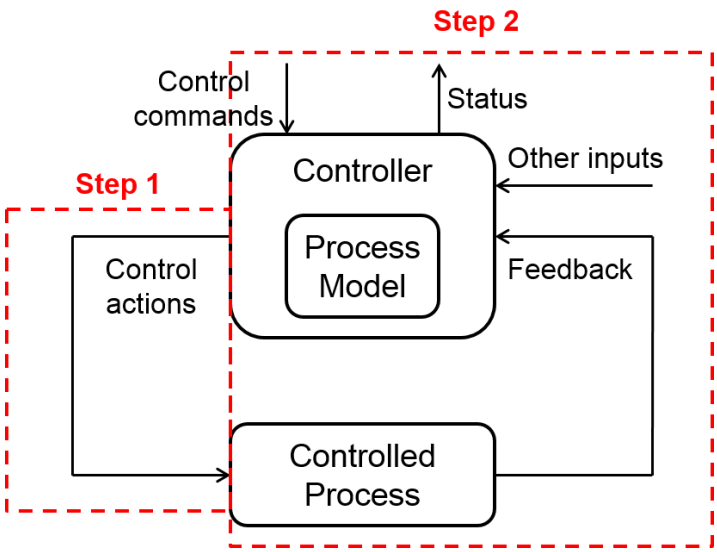

Figure 7: Generic functional control diagram

Step 1

In order to support step 1, “Unsafe control action identification”, a procedure has already been developed. This procedure is based on three independent parts and considers that many control actions are only hazardous in certain contexts [12].

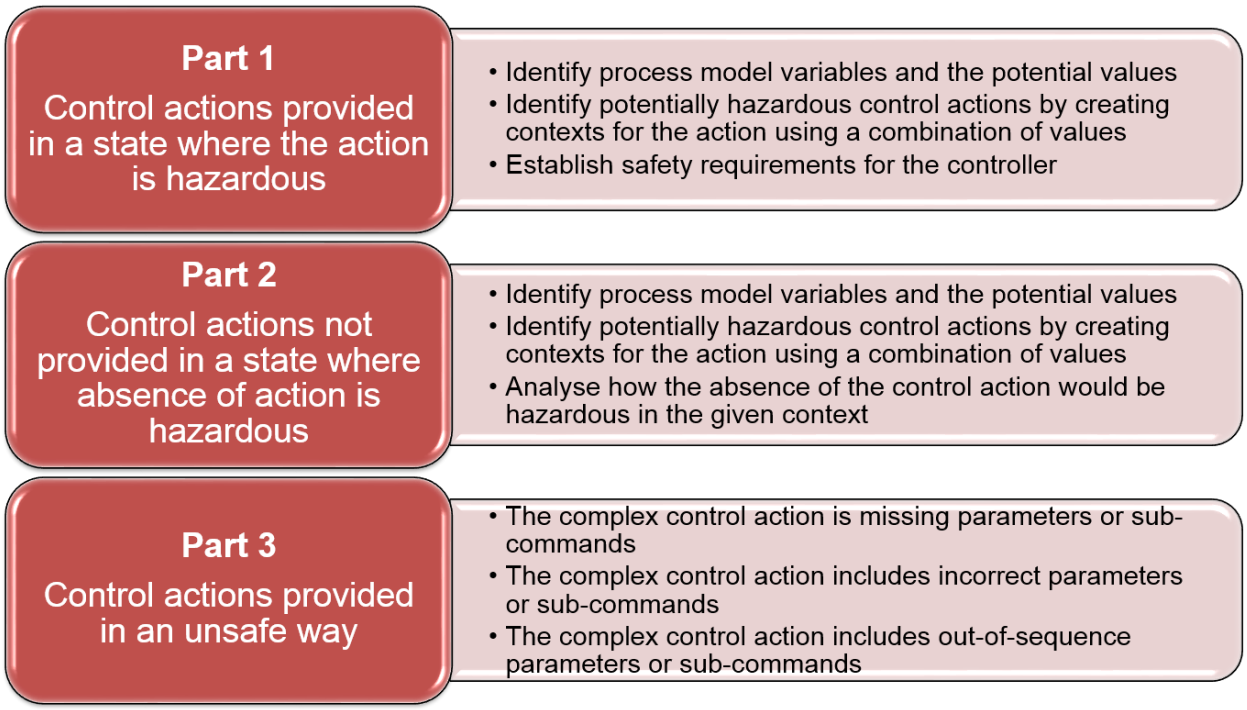

Figure 8: Step 1 – Identification of unsafe control actions

Part 3 is targeted at complex commands that can be provided in more than one way. Researchers are still focusing on this point and on whether there are alternatives to this approach.

At the end of Step 1, it would be possible to translate the identified hazards into safety constraints that must be enforced by the design of the system. Step 2 will identify the scenarios that can lead to those hazards.

Step 2

During this step, each control loop in the safety control structure is analysed in order to identify the scenarios (causal factors) that can lead to the unsafe control actions identified in step 1. Figure 9 presents a generic control loop structure.

Once again, after Step 2 is completed, any identified causes that lead to unsafe control actions must be eliminated or controlled during the design phase.

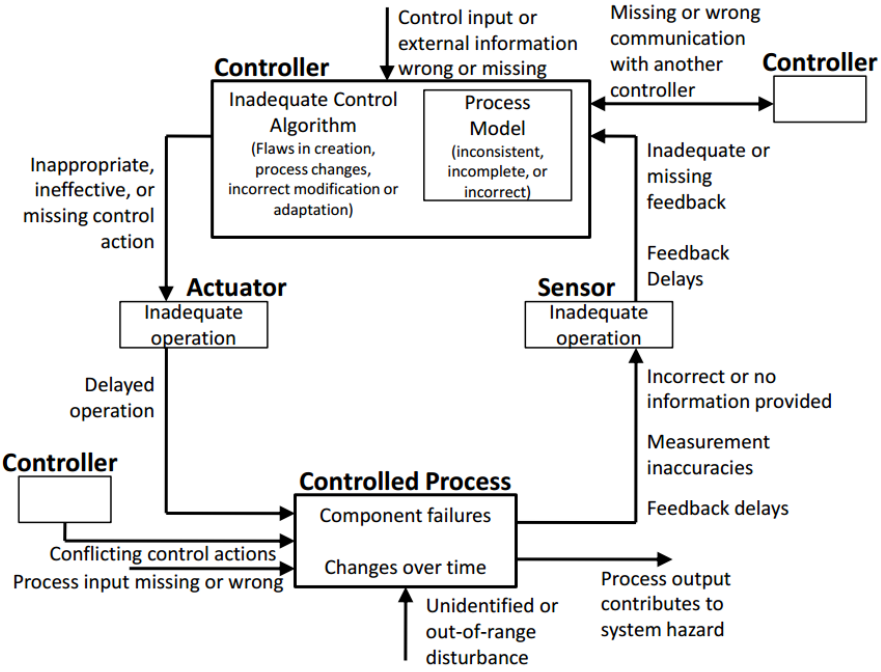

Figure 9: Step 2 – Identifying the causes of unsafe control actions [13]

Case studies

STAMP has been used in several case studies in different domains, including in the rail industry, such as:

- **Re-analysis of a high-speed train accident in China that took place on July 23rd, 2011, where 40 people were killed and 120 injured on the Yong-Wen High Speed Line [8].**
- Comair 5191 LEX Accident [9];
- JAXA HTV [10];
- Infringement of separation between B738 & A319, Area Control Centre – Blue [11];

CONCLUSION

The role of human factors has always been considered a key element in risk assessment. Several techniques have been employed to quantify its influence in situations that lead to accidents, some of these techniques use a quantitative approach and others a qualitative one. However, none of them have proven to be effective in complex human behaviour modeling and, as the systems are becoming more and more complex, the need for considering the connection between human performance, human culture and external factors which have impact on human behaviour increases significantly.

The present paper demonstrates that the STAMP model is able to access the system as a whole, focusing on handling the dynamic nature of complex socio-technical systems in order to identify incidences of inadequate control that can lead to an accident. Some approaches have already been applied in the railway industry in order to develop some adaptations of the STAMP model that can better meet the specific requirements of the railway industry. It is important that all railway partners contribute, through the sharing of previous experience, tacit knowledge and new ideas, in order to improve and define new procedures to effectively implement the STAMP model. This will provide additional guidance to identify, with a high level of accuracy, the unsafe control actions

that could lead to a hazardous situation and the scenarios wherein these unsafe control actions can occur. Additionally, CENELEC standards should include the application of these methodologies as a requirement for risk evaluation.

Future work / next steps:

- Application of a tailored STAMP methodology to real cases;
- Tools adaptation or new tools development to support the technique;
- Standards improvement. Critical Software is currently involved in the European FP7 CECRIS project, particularly in defining gaps relating to the disciplines of human factor and security.

It is the authors' belief that if the railway industry adopts these new hazard and risk assessment methodologies, the accident numbers presented in this paper as a challenge to overcome will decrease.

REFERENCES

[1] "Human Reliability and error in Transportation systems", B. S. Dhillon, 2007

[2] "International Encyclopedia of Ergonomics and Human Factors", E. Hollnagel, 2006

[3] "Understanding the human factors contribution to railway accidents and incidents in Australia. Accident Analysis and Prevention40", M. T. Baysari, A. S. McIntosh, J. R. Wilson, (2008) 1750–1757.

[4] SHOOMAN ML, Avionics software problem occurrence rates, pages 53-64, IEEE Computer Society Press, 1996

[5] An Overview of Human Reliability Analysis Techniques in Manufacturing Operations, Valentina Di Pasquale, Raffaele Iannone, Salvatore Miranda and Stefano Riemma

[6] A Review of Accident Modeling Approaches for Complex Socio-Technical Systems; Zahid H.Qureshi; University of South Australia

[7] A New Accident Model for Engineering Safer Systems; Leveson, N.; Safety Science; 2004

[8] A STAMP Analysis on the China-Yongwen Railway Accident; Song, T., Zhong, D., Zhong, H.; Beihang University; 2012

[9] A STAMP Case Study - Comair 5192 LEX Accident; Paul Sidney Nelson; Lund University, Sweden

[10] Modeling and Hazard Analysis using STPA; Takuto Ishimatsu, Nancy Leveson, John Thomas, Masa Katahira, Yuko Miyamoto, Haruka Nakao; MIT and JAXA/JAMSS

[11] A Qualitative Comparative Analysis of SOAM and STAMP in ATM Occurrence Investigation; Richard Arnold; Lund University, Sweden

[12] Performing Hazard Analysis on Complex, Software-and Human-Intensive Systems; J. Thomas, S.M.; N. G. Leveson, Ph.D.

[13] STAMP experienced users tutorial, J.Thomas, B. Antoine, C. Fleming, M. Spencer, Q. Hommes, T. Ishimatsu, J. Helferich

[14] A new method for human reliability assessment in railway transport, D. Schwencke, T. Lindner, B. Milius, M. Arenius, O. Strater, K. Lemmer

[15] The Human factor in risk assessment: Methodological comparison between human reliability analysis techniques

[16] The Human Factors Analysis and Classification System–HFACS, Scott A. Shappell, Douglas A. Wiegmann