



HAL
open science

Performance shaping factor based human reliability assessment using valuation-based systems: application to railway operations

Subeer Rangra

► **To cite this version:**

Subeer Rangra. Performance shaping factor based human reliability assessment using valuation-based systems: application to railway operations. Human-Computer Interaction [cs.HC]. Université de Technologie de Compiègne, 2017. English. NNT: 2017COMP2375 . tel-02004305

HAL Id: tel-02004305

<https://theses.hal.science/tel-02004305>

Submitted on 1 Feb 2019

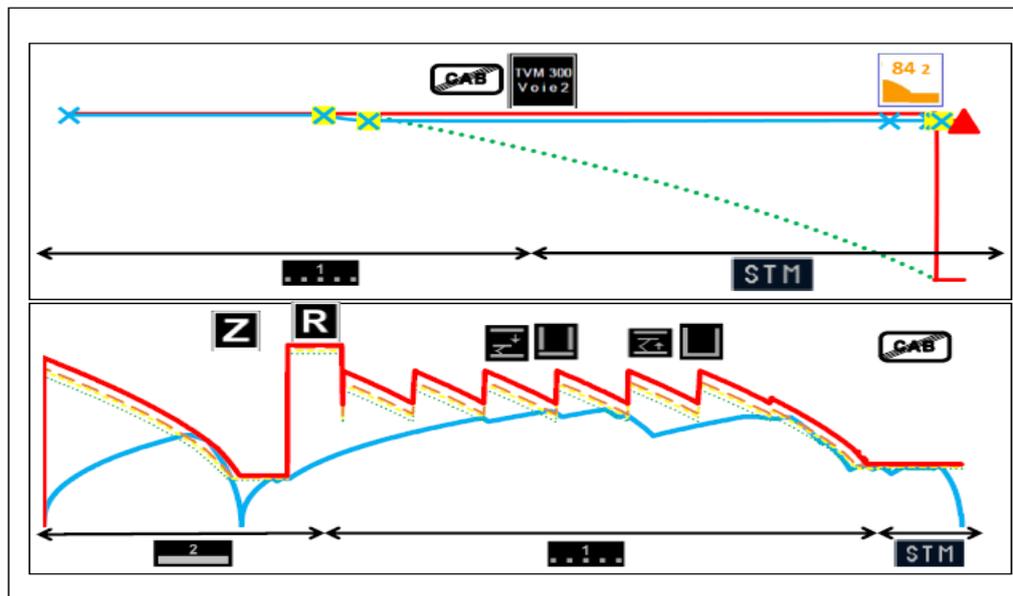
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Par **Subeer RANGRA**

Performance shaping factor based human reliability assessment using valuation-based systems : application to railway operations

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC



Soutenue le 3 octobre 2017

Spécialité : Technologies de l'Information et des Systèmes :
Unité de recherche Heudyasic (UMR-7253)

D2375

Performance shaping factor based human reliability assessment using valuation-based systems – application to railway operations

Subeer RANGRA

Spécialité : Technologies de l'Information et des Systèmes

Thesis defended on 3rd October 2017, before the jury composed of:

Jury president:

Véronique Berge-Cherfaoui

professeur des universités, Université de Technologie de Compiègne, Heudiasyc laboratory

Jury members:

Simon COLLART-DUTILLEUL

directeur de recherche, Université Lille Nord de France, IFSTTAR (reviewer)

Philippe WEBER

professeur des universités, University of Lorraine, CRAN Laboratory (reviewer)

Julie BEUGIN

chargée de recherche, Université Lille Nord de France, IFSTTAR

Jérémie GUIOCHET

maître de conférences (HDR), Université de Toulouse, LAAS laboratory

Mohamed SALLAK

maître de conférences (HDR), Université de Technologie de Compiègne, Heudiasyc laboratory (thesis supervisor)

Walter SCHÖN

professeur des universités, Université de Technologie de Compiègne, Heudiasyc laboratory (thesis supervisor)

Frédéric VANDERHAEGEN

professeur des universités, Université de Valenciennes, LAMIH laboratory (thesis supervisor)

Université de Technologie de Compiègne

Heudiasyc UMR CNRS 7253

Labex MS2T



Standing on the shoulder of giants.

*To my family. Loving mother, supportive
father, and driven sister.*

Acknowledgements

I would like to begin by extending a heartfelt gratitude to each one of my three supervisors Mohamed Sallak, Walter Schön and Frédéric Vanderhaegen. For these three years of guidance, perseverance, and valuable time; the constant support and encouragement for this thesis work and other activities. The endless stream of ideas, freedom, and critiques, which refined not only this work, but also me as a person.

I am grateful to the jury members for doing me the honors to evaluate my work. The reviewers for their valuable remarks and critiques of this manuscript. All the jury members for their valuable questions and comments; and undoubtedly their time. In this final version of the thesis I have tried to address most of your questions and remarks.

Thanks to the various people and entities of Heudiasyc lab., and UTC. First and foremost Ali Charara, whose supportive and principled nature towards PhD students, was both greatly appreciated and a learning experience for me. I got to witness firsthand the supportive work environment that the Heudiasyc lab was, for a PhD student, and how fortunate I was to be a part of it. I would like to thank the whole administrative team at lab (the famous 2nd floor!) – the secretaries, support staff and others, who were some of the most caring and helpful people. Thanks also the various entities at the level of UTC. I would also personally like to thank the funding framework of Labex MS2T, all the people and the entities involved: the administrative and the direction staff who ensured both financial and scientific support to this work. A small word of thanks is also owed to the research engineer at the Plateforme ferrovaire.

I also experienced the transition from studying to working, in a very conducive and particular work environment – a research lab. This transition was all the more agreeable and due to the friends and colleagues. Their time, the enriching (or even nonsensical) discussions, company during late nights and weekends in the lab and outside for endless coffee (or other more animating beverages) sessions, all made me a better person. For these three years I would like to thank: Neeraj, Shriram, Ayyoub, Freddy, Youcef, Hafida, Idir, Sagnik, Nicola, Ahmet, and may others who I

inadvertently forgot to name.

This is being written around one month after this manuscript was finished. In retrospect there are probably a significant number of things that I would have liked to write, explain or visualize for the author differently.

I am sure there are many other people and entities, who I might have forgotten to mention here. Rest assured their support, implicit or explicit is something I am grateful for.

Some acknowledgment is also owed to the tools and communities which made this work relatively easier: \LaTeX (taking into account the “new” problems it made me address), Mendeley, Evernote, Office, Matlab, Docear, and others.

Last but not least I would like to partially dedicate this work to my family, my mother whose unending love for her son kept him going, my father whose unconditional support kept him on track, and my sister who for all her drive in doing what she wants and sees right, taught me a thing or two.

This thesis is partially dedicated to the research and engineering community, the reference list of this manuscript is a testimony to their groundwork upon which this work is built. “*If I have seen further it is by standing on the shoulders of Giants.*”
– Isaac Newton.

Contents

Acknowledgements

Contents	i
List of figures	v
List of tables	ix
Acronyms	1
1 General Introduction	5
2 Theoretical background	13
2.1 Reliability and risk analysis basic notions	14
2.1.1 Some reliability and risk assessment methods	15
2.1.1.1 Fault trees	16
2.1.1.2 Failure modes, effects and criticality analysis . . .	16
2.1.1.3 Event tree	16
2.1.1.4 Bayesian networks	17
2.1.1.5 Valuation-based systems	18
2.2 Mathematical framework to represent uncertainty	20
2.2.1 Belief Functions Theory	21
2.2.2 Some BFT-based combination rules	24
2.2.2.1 Dempster's rule	25
2.2.2.2 Yager's rule	25
2.2.3 Comparison between BPAs using interval and distance metrics	26
2.3 Conclusions	27
3 Problem background	29
3.1 Labex MS2T: control of technological Systems-of-Systems	30
3.2 Transportation system-of-systems and safety	31
3.3 Human errors: some notions	34

3.3.1	Some perspectives on analyzing human errors	34
3.3.2	Human error: classification and taxonomies	36
3.3.2.1	Person approach	37
3.3.2.2	Systems approach	38
3.4	Application context: railway operations	38
3.4.1	European Rail Traffic Management System	39
3.4.1.1	ERTMS application levels	40
3.4.1.2	ERTMS/ETCS braking curves and train driver DMI	41
3.4.2	European regulations and human errors	45
3.5	Conclusions	47
4	A survey on human error quantification and its application to railway transportation	49
4.1	Introduction	51
4.2	Quantitative HRA – variables, data and frameworks	54
4.2.1	Identification of the variables of a quantitative model	56
4.2.1.1	Human failure event identification	57
4.2.1.2	Performance Shaping Factors	57
4.2.2	Error modeling and sources of data	62
4.2.2.1	Expert data	62
4.2.2.2	Empirical and experimental data	64
4.2.3	Quantification frameworks: a classification	65
4.2.3.1	Multiplier-like	66
4.2.3.2	Expert focused	70
4.2.3.3	Probabilistic graphical model-based	74
4.2.4	A summary and comparison	78
4.3	HRA in rail transportation	81
4.3.1	Variables of a quantitative framework for railway	82
4.3.1.1	Human failure event for rail operation	82
4.3.1.2	PSFs for railway and related works	85
4.3.2	Some quantitative considerations of human errors and frameworks in railway	91
4.4	Discussion towards a complete railway HRA methodology	94
4.5	Conclusions	96
5	PRELUDE: Performance shaping factor based human reliability assessment using valuation-based systems	99
5.1	The PRELUDE methodology	100

5.1.1	Qualitative part	101
5.1.1.1	Performance Shaping Factor list and evaluation . .	102
5.1.1.2	Identification of HFEs and safety critical context .	103
5.1.2	Quantitative part	106
5.1.2.1	The expert elicitation process	107
5.1.2.2	Combination of expert data	109
5.1.2.3	Transformation	112
5.1.3	Quantification and sensitivity analysis	116
5.1.3.1	Assigning the direct evidence and quantification .	116
5.1.3.2	Sensitivity analysis	117
5.2	Case study	121
5.2.1	Step 1. Qualitative part: HFE and PSF(s) identification . . .	121
5.2.2	Step 2. Quantitative part: Expert elicitation, data combina- tion and transformation	125
5.2.3	Step 3. Quantification data and results	128
5.3	Conclusions	133
6	Feasibility study of PRELUDE with data from simulator experimenta- tion	137
6.1	Introduction	139
6.2	Experimental protocol using an operational simulator to obtain human reliability data	140
6.2.1	Simulator set-up: description of the ERTMS operational Sim- ulator	141
6.2.2	The simulation environment	143
6.2.2.1	Rail track and procedures	144
6.2.2.2	Description of scenario runs	144
6.2.2.3	Explanation and basic training	148
6.2.3	Output data sources and analysis	149
6.2.3.1	Objective data: source and calculation of scores . .	149
6.2.3.2	Subjective data: source	154
6.2.4	Experimental campaign and preliminary discussion of the collected data	155
6.2.4.1	Objective data	157
6.2.4.2	Subjective data	158
6.3	Using data obtained from experimental protocol in PRELUDE . . .	163
6.3.1	Pre-analysis: A classification of subjects	163

6.3.2	Objective data usage: combination with expert data – input to PRELUDE Quantitative part	166
6.3.3	Subjective data usage: Retrospective identification of PSF self estimation – input to PRELUDE qualitative part	172
6.4	Discussion	176
6.5	Conclusion	178
7	Conclusions and Perspectives	181
A	Appendix	187
A.1	Appendix to the PRELUDE methodology	187
A.1.1	The flowchart representation of the PRELUDE methodology	187
A.1.2	Case study: extraction of data from the accident report . . .	189
A.2	Experimental campaign questionnaires, source code and additional data collected	193
A.2.1	NASA Task Load Index (NASA-TLX) Questionnaire	193
A.2.2	Subjective questionnaires: Pre, post and PSF subjective questionnaire	195
A.2.3	Code used to extract data from Eurocab.log	203
A.2.4	Subjective questionnaires data - Pre, post questionnaire . .	208
	References	211

List of figures

2.1	A classic risk matrix for determining the severity of a hazard for the system under scrutiny [CENELEC, 1999]	14
2.2	An example of Event tree analysis: a gas pipe carrying a flammable gas	17
2.3	BNs of a simple structure representing relation between three variables	18
2.4	A small example and the associated variables modeled as a VBS (a VN)	20
3.1	An overview of <i>Labex MS2T</i> and the positioning of this thesis.	30
3.2	A Swiss cheese model representing accident causation, from [Reason, 2000]	37
3.3	Unsafe acts or human failures breakdown in a person approach to analyze a human error	37
3.4	Architecture of an ERTMS/ETCS level 1. Source: [Wikipedia, 2017] .	40
3.5	Architecture of an ERTMS/ETCS level 2. Source: [Wikipedia, 2017] .	41
3.6	Architecture of an ERTMS/ETCS level 3. Source: [Wikipedia, 2017] .	41
3.7	Overview of the ETCS braking curve and its related supervision limits. Source: [European Railway Agency, 2016]	42
3.8	ETCS DMI's speed displays different representations based on the train speed vs. braking curves (colors represent the different sped curves also indicated by the labels on the right.)	43
3.9	Speed Curve parameters defined as ETCS fixed values Source: [UNISIG, 2012] (colors represent the different sped curves also indicated by the labels on the right.)	44
3.10	Factors which can have an impact on RAMS of railways as per EN50126 [CENELEC, 1999]	46
4.1	Steps of a HRA quantitative framework	56
4.2	Example of a PSF list: As proposed by the CREAM model [Hollnagel, 1998], and their states	59
4.3	Example of a PSF list: SPAR-H's PSFs [Gertman et al., 2005] and the corresponding levels.	60
4.4	SPAR-H PSF states and respective multiplier values assigned for a state of a PSF for an action task[Gertman et al., 2005]	68

4.5 SPAR-H quantification steps (boxes in green show the steps performed in SPAR-H)	69
4.6 An excerpt from the tabular predefined values given with the HEART methodology [Williams, 1986]: GTT's <i>NHEP</i> values (left), and EPC multipliers (right).	70
4.7 Integration of ATHEANA's results in a PRA model using two possible ways, adapted from [J. Forester et al., 2007]	72
4.8 ATHEANA quantification steps (boxes in green show the steps performed in the ATHEANA methodology; half green-red shows that it is not performed in ATHEANA, but guidelines are provided)	73
4.9 A BBN model structure proposed in [Groth and Swiler, 2013] as the interpretation of SPAR-H's quantitative model	75
4.10 Quantification steps of a probabilistic graphic models HRA quantification framework	78
4.11 Number of responses from railway entities on "how often they use specific human factors techniques?" taken from [Kecklund et al., 2013]	82
4.12 Conceptual Fault Tree for the functional analysis of the ETCS (application level 2) within an operational railway environment [UNISIG (Union of Signalling Industry), 2016] and the FMEA for a <i>driver error</i> base event.	84
4.13 Functional decomposition of railway system, to identify the task to be performed and associated behavior (an HFE's task characteristics), adapted from [Vanderhaegen, 2001]	85
4.14 ERA study data organization and exaction of PSF. The parts of the study which can be used for the PSF list are colored in red.	88
4.15 Predefined values for human error (HEP), as used in some rail applications	92
5.1 Overview of the PRELUDE methodology. Step 1. is the qualitative part which aims to identify and characterize a safety critical context, as HFE and a set of PSFs; Step 2. is the quantitative part, which builds the VBS model from expert data; and Step 3. presents the final HFE quantification and sensitivity analysis results.	101
5.2 The usage of EUAR HF study to identify the safety critical context, a refinement of the context related to an human function and in turn an HFE, is done using the data as marked in red.	105
5.3 Graphical representation of the example HFE's implementation in VBS.	119
5.4 Sensitivity analysis results for example HFE and associated PSFs	120

5.5	Accident scenario: speed of the train vs. distance from the point of accident. The data points (cross marks) are other events as identified in the investigation report and the HFEs. Also a time scale is given to represent the time duration of the analyzed scenario	122
5.6	Expert data (first three bars) and data obtained from the combination rules, for the case study.	127
5.7	VBS model of HFE1 and its HFTC.	129
5.8	VBS model of HFE2 and its HFTC.	129
5.9	VBS model of HFE3 and its HFTC.	129
5.10	Lower and upper bounds for an HFE's true state, different models are built for when combining expert elicitation using average (A), weighted average (WA), vote/independent consensus (V), Dempster's (D) and Yager's (Y) combination rules	131
5.11	Sensitivity analysis results of PSFs for the context of <i>HFE2</i> and <i>HFE3</i>	133
6.1	Overview of PRELUDE's methodology with the data requirements in dashed boxes, and experimental protocol's inputs and outputs	140
6.2	Overview of the experimental protocol showing the inputs: the PSF list from PRELUDE, and the outputs: the objective and subjective data	141
6.3	ERTMS/ETCS Operational Simulator set-up: a picture, and its architecture listing the functions and different machines	143
6.4	ETCS on-board generated speed profile of the track used in present work	144
6.5	Speed curve score calculation - in purple shaded area over EBI curve, that is the braking curve for the Warning component V_{EBI}	151
6.6	Modified dV parameters to take into account speed difference from indicated speed for score calculation	152
6.7	The experimental campaign with some photos of subjects during the course of a session	156
6.8	Box plot of the <i>Safety score</i> of all the subjects for each run	157
6.9	Box plot of the <i>Time score</i> of all the subjects for each run	157
6.10	Box plot of the TLX global score	158
6.11	Average TLX sub-scale scores for all the parameters of TLX: ratings and weighted values	160
6.12	The number of responses to PSF subjective questionnaires, "poor" and "I am not sure", for all the runs	162
6.13	Classification of subjects into groups based on their average scores .	164
6.14	Box plot of safety scores for subjects in group 1 (top left), group 1 (top right) group 3 (bottom)	165

6.15	Box plot of time scores for subjects in group 1 (top left), group 1 (top right) group 3 (bottom)	166
6.16	Overview of the usage of objective data from simulator experimentation in the PRELUDE's quantitative part	167
6.17	Expert data and simulator data (p_{SS}^{G3}) value, and their combination using: average, weighted average (top), Dempster's and Yager's rule (bottom), line plot shows the conflict value ($1 - k$) on the secondary x axis.	170
6.18	Quantification of the HFE from expert (only) data combination (Average, Weighted Average, Vote, Dempster, Yager's), and combination of average combined expert data (A) with the average combined simulator data bar plot ($A + SA$)	172
6.19	Overview of the usage of subjective data from simulator experimentation in the PRELUDE's qualitative part	173
6.20	Self estimation level of the subjects using average values of their p_{SS} (the higher the worse for safety), and performance score of TLX, with the three groups of subjects as identified in subsection 6.3.1	175
A.1	A flowchart description of the PRELUDE methodology.	188

List of tables

3.1 Fixed values data and explanation for ETCS braking curves, also visible in Figure 3.9	44
4.1 An example of definition of a relational data between the states of PSF and state of an HFE	77
4.2 Some advantages and disadvantages of the previously discussed three classes of quantitative HRA frameworks	79
4.3 A comparison of some main HRA methodologies	80
4.4 PSF list with considered definitions and quantification levels, adapted from [Rangra et al., 2015a]	90
5.1 PSF list with considered definitions and levels [Rangra et al., 2015a]	104
5.2 Defining Human Failure Type Context for the example HFE	107
5.3 The HFE’s description and relevant context description and question statements	110
5.4 Combination rules and their hypothesis	111
5.5 Direct belief structures for HFE: from the R-PSF equivalent as identified focal sets and associated <i>bpa</i> values.	119
5.6 Marginalization results for example HFE on Ω_{HFE}	119
5.7 HFE and relevant procedures from the accident investigation report and national regulations	123
5.8 Identification of PSFs for defining HFTC for HFEs from the accident scenario	124
5.9 Context description and questions for HFE2 sent to the experts	126
5.10 Direct belief structures (focal set and <i>bpa</i>) for the PSFs in the case study, identified similar to as in Table 5.5 from the R-PSF equivalent	129
5.11 Top: <i>Middle of the probability interval</i> for the variable <i>HFE1</i> value of interest (<i>true</i>) as obtained after combining data using different combination rules. Bottom: The pairwise <i>distance metric</i> d_J from equation 2.7, between the <i>bpas</i> obtained for <i>HFE1</i>	130
6.1 Normalized weights for ETCS reference speeds	152
6.2 Expert combined data and data from experimentation	169

6.3 Self estimation levels based on Safety score and subjective performance ratings	174
A.1 NASA TLX descriptors: questions and scale	194

Acronyms

ADAS: Advanced Driver Assistance Systems

ATHEANA: A Technique for Human Error Analysis

ATO: Automatic Train Operation

ATP: Automatic Train Protection

AWS: Automatic Warning Systems

BBN: Bayesian Belief Networks

BFT: Belief Functions Theory

BNN: Bayesian Belief Networks (BBN)

BPA: Basic Probability Assignments

BPA/*bpa*: basic probability assignments

CENELEC: European Committee for Electrotechnical Standardization

CPT: Conditional Probability Tables

CREAM: Cognitive Reliability and Error Analysis Method

CSM: Common Safety Methods

EBI: Emergency Brake Intervention

EC: Error Contexts

EN: Evidential Network

EOA: End of Authority

ERA: Event Trees Analysis

ERTMS: European Rail Traffic Management System

ETCS: European Train Control System

EUAR: European Union Agency for Railway (old ERA: European Railway Agency)

F/Nf: Franchissable / Non Franchissable (Passable and non Passable)

FME(C)A: Failure Modes Effects (and Criticality) Analysis

FTA: Fault Tree Analysis

GSM-R: GSM mobile communications standard for Railway

GTT: Generic Task Type

HEART: Human Error Assessment and Reduction Technique

HEP: Human Error Probability

HERA: Human Event Repository and Analysis

HFE: Human Failure Event

HFTC: Human Failure Type Context

HMI: Human Machine Interface

HRA: Human Reliability Analysis

HSI: Human System Interface

HuPeROI: Human Performance Railway Operational Index

IM: Infrastructure Manager

LRBG: Last read balise group

MERMOS: French for Assessment Method for the Performance of Safety Operation)

NARA: Nuclear Action Reliability Assessment

NASA-TLX: National Aeronautic and Space Administration-Task Load Index

NSA: National Safety Authority

PIF: Performance Influencing Factors

PRA: Probabilistic Risk Assessment

PRELUDE: Performance shaping factor based human reliability assessment using valuation-based systems

PSF: Performance Shaping Factor

RAMS: Reliability, Availability, Maintainability, and Safety.

RARA: Rail Action Reliability Assessment

RBS: Radio block center

RU: Railway undertakings

SBI: Service Brake Intervention

SPAR-H: Standardized Plant Analysis Risk Model – Human reliability analysis

SRS: System Requirement Specifications

SoS: System of systems

THERP: Technique for Human Error Rate Prediction

TRACER: The technique for the retrospective and predictive analysis of cognitive errors

VBS: Valuation-Based Systems

VN: Valuation Networks

General Introduction

Humans are and will remain one of the critical constituents of a sociotechnical system. It has been widely reported in rail and road transportation that majority of accidents are caused at least in part, by some form of human error [Evans, 2011] [Kyriakidis et al., 2015b]. However, increasing complexity of such systems makes it difficult to identify the reliability of the subsystems (including a human) and inversely the system. A System-of-Systems (SoS) view provides with adequate directions to handle this problem [Rangra et al., 2015b]. When we consider human controller as a component of the system it exhibits autonomy, operational independence and induces emergent properties [Wilson, 2014], co-operating with other components towards a common goal. On the other hand, the systems approach to human error states that “*humans are fallible and errors are to be expected, even in the best organizations*” [Reason, 2000]. Further, these errors are the consequence of inadequate conditions residing within complex systems. Such an approach is more recognized, and used in a retrospective analysis, i.e. accident analysis, and forms the basis of so-called Systemic Accident Analysis (SAA) approaches [Underwood and Waterson, 2013] [Leveson, 2015].

However, previous works for the railway domain are for the most part qualitative. Further, such an analysis is not compatible with quantitative analysis of technical failures which is an essential part of an integrated safety and risk analysis. In addition, risk analysis related to human interactions and their evaluation, need to evolve and be recognized by regulatory and operational authorities; as evident by various human factor and risk analysis studies carried in the last few years by the European Union Agency for Railways (EUAR) [Det Norske Veritas, 2010] [Kecklund et al., 2013] [Pickup et al., 2013]. At the regulatory level, risks related to human errors and their assessment need further research, as recognized in the latest amendment to Common Safety Methods (CSM) [European Railway Agency, 2015a]. The study [Kecklund et al., 2013] of rail entities, concluded that there is a “*need to increase the knowledge on risk assessment of human interaction within the European railway system and to further increase the exchange of information on this*

topic within the European railway community”.

The family of methods called Human Reliability Analysis (HRA) aim to systematically integrate the risk associated with human interactions for system-level risk analysis. Since the early 70's first generation HRA methods have been developed with broadly similar features, e.g. task analysis, nominal and modified probabilities for human failures, etc. The second generation aimed for a less of a focus on individual errors, and more on determining the factors and conditions around said errors; some focused on cognitive model-based methods aiming for a complete capture of human performance. Nevertheless, such classification is often not sufficient, and a clear identification of desired, and valid techniques is not straightforward. Some newer methods are classified as first generations, some second generation methods have been said to be too costly to implement, so much so that first generation are often preferred. Thus, work is needed to identify the good-practices.

As a starting point, human reliability and human error can be defined in terms of the causes of human behavioral dysfunction and/or their consequences for the system. Most HRA methods are thus, risk assessment-based and/or cognitive model-based methods. These assess or analyze the risks of human or system dysfunction due to human actions, which are evaluated in relation to the causes of human behavioral dysfunction. The causes, or in general a performance-degrading context is characterized as Performance Shaping Factors (PSFs). Towards risk analysis objectives, these factors affect human performance, and in-turn system safety. More specifically, PSFs allow the consideration of human's own characteristics along with environment which affect human performance in a negative or positive manner [Blackman et al., 2008]. The impact of PSFs can be assessed through different criteria such as safety, production of services, task load, stress, attention, etc. by focusing on a multi-criterion consequence analysis of PSFs and on their interdependencies [Vanderhaegen, 2001] [Vanderhaegen, 2010]. An HRA model, in their simpler forms models the relation between PSFs and human performance. Here the performance is related to a system safety criteria, to evaluate system-level risk.

Furthermore, most of the work in the domain of HRA is done in, and for the nuclear domain. Over the years most PSFs sets have gone through multiple revisions and critiques giving them a refined definition, and hierarchical structuring among other classifications. An extensive list of PSF is advantageous particularly in performing a complete and detailed analysis, e.g. when doing HRA in design phase, qualitative analysis to pinpoint exact causes of errors, etc. Owing to different operational context and functional needs from human operators, such exhaustive

lists need to be modified, significantly if not completely to account for domain-specific considerations. Further, accident reports can be used to identify PSFs responsible or involved in most of the accidents. Additionally, if a quantitative analysis is desired, the correlations between factors have to be identified in order to simplify their integration into a human reliability assessment.

When designing a SoS, a quantitative risk analysis identifies undesirable scenarios for which the designers have to specify material barriers or procedures in order to make them acceptable, and reduce the residual risk level under a threshold (a risk acceptance criteria). This process does not consider that the human operators can sometimes remove some of these barriers in order to optimize the compromise between performance criteria such as safety, task load, quality or production of service for instance [Sedki et al., 2013] [Vanderhaegen et al., 2011]. The risk assessment of barrier removals is an challenging topic, and requires a strong collection of field data to develop relevant human behavioral models. For instance, models based on dissonance engineering can support the representation of rule or knowledge of a SoS functioning and use, and can identify possibly dangerous or beneficial dissonances involving human, technical, environmental or organizational factors [Vanderhaegen and Carsten, 2017] [Vanderhaegen and Zieba, 2014] [Qiu et al., 2017].

Quantitative human reliability, like most reliability analysis problems, although maybe more severely suffers from lack of data problem. When working with a lack of empirical data, expert elicitation, conditional data, prior probabilities and data combination are often employed. The use of probabilistic graphical models is an interesting framework for HRA application. These models not only allow modeling causal effects of factors, but also allow using different sources and types of data. Thus, usage of Bayesian networks in the domain of reliability, and particularly HRA has seen growth in recent years [Mkrтчyan et al., 2015]. To this extent, such frameworks are particularly easy to use in interaction with experts, and are mathematically more expressive than traditional approaches. These expert systems, not only allow a standard-form model which can be used by analysts, but also are helpful in improving the transparency and repeatability of assessments.

For the quantification of rare events (such as human failures) managing uncertainty in data is an important and challenging task. Towards the objectives of an accurate representation, and subsequent evaluation, it is frequently classified by its source. The one originating from natural randomness is called aleatory, and the one originating from a lack of information is termed epistemic. To address this task appropriate uncertainty representation and management is often desired. Various mathematical frameworks deal with such problems. Dempster-Shafer theory

also known as evidence theory or belief functions theory (BFT) allows usage of upper and lower probability instead of precise values with generalization of the Bayesian theory of subjective probabilities. The main contents of this theory are the combination and representation of evidence or knowledge. A model based on BFT can be represented with a Valuation-Based Systems (VBS). VBS was proposed as a general language for incorporating uncertainty in expert systems. Such evidential networks present some notable qualities when dealing with uncertainty and decision making. Thus, this framework offers adequate tools to work with uncertainty in data and experts. VBS can also be used to represent several domains for combination of the information such as Bayesian probability theory, possibility theory, BFT, propositional logic, etc. A VBS-BFT framework can represent and propagate both (epistemic and aleatory) types of uncertainty. And it is able to integrate all types of sources (accidents, incidents, experts and simulators) of data to build a robust HRA model. Thus for a quantitative HRA modeling, a framework of VBS implementing BFT presents adequate mathematical tool set.

Ideal case for HRA remains to have extensive experimental campaigns to obtain robust data-set using simulator trials. Simulators allow obtaining objective *human performance* data using objective criteria (success, failure, etc.) and subjective data using standard feedback questionnaires. To characterize effects of contexts on objective performance, nominal human error probability calculation, etc. Such methodology can be applied to already existing training programs for other domains. Often used in training, such simulators can be a valuable source of quantitative HRA data, which can be used to model, verify, and validate, and to respond to why performance was inadequate. Here, subjective questionnaire present another dimension to interpret human performance. They support interpretations that are not obvious only from objective data. The use of multidimensional, subjective tools like NASA TLX (Task Load Index) is merited and allows getting a complete picture from an operator's point of view. More than the objective measures, a subjective component to a HRA or safety related activity is crucial, towards making improvements or a simple formalized feedback.

Thus, this work focuses on the needs of railway domain. It is a new, generic framework inspired from current HRA practices, and aims to address some issues of the HRA methods as discussed above. To arrive at an quantitative HRA methodology for railway application, a critical survey on human error quantification techniques was performed. The main contribution of this thesis is a new original and generic framework of human reliability analysis (HRA) applied to the railway domain [Rangra et al., 2017b]. This complete qualitative and quantitative HRA methodology is called PRELUDE (acronym for a Performance shaping factor based

human Reliability assessment using valuation-based systems). It aims to address aspects of operational risk analysis of a rail operator's activities. The qualitative part characterizes situation or operational context to identify safety critical performance shaping factors (PSFs). These PSFs are identified from domain factors studies, accidents statistics, and PSF lists in-general, which are often analyzed in various HRA models [Rangra et al., 2015a]. More precisely, to respond to the question of what PSFs to consider for an HRA, three point of views are considered. First, data from past accidents – what PSFs are implicated in human error related accidents, second, an operational safety-oriented analysis of human functions and goals in rail operation, and third lessons from general HRA methods – as to what PSFs are used in other HRA models irrespective of the application domain.

Our quantitative proposition is a framework of VBS, and the BFT as the underlying mathematical framework. In this part, it is used to build an expert system using human reliability data from the domain experts. Multiple experts are elicited on human reliability data, the data is combined using BFT-based combination rules to manage, in particular conflicting opinions and lack of information. This combined expert data is then transformed to build the VBS. This transformation proposal is a formal framework to build a human reliability model in VBS from conditional expert data. The VBS model thus built, provides decision-making using probability intervals, quantifying a human failure event given an operational context. Sensitivity analysis is used to establish a priority ranking among the PSFs.

Finally, a case study of a real high-speed railway accident scenario is presented to demonstrate the PRELUDE's usage for a retrospective analysis. The focus is the train driver, with the operation context being a section of a high speed railway line, with appropriate signaling. Qualitative data on the scenario (Human Failure Events, PSFs, etc.) are identified from the accident investigation report, regulatory and operational reference documents. Domain experts were elicited, their data combined, VBS models built and human error probability was obtained. In the identified scenario and contextual data it was able to identify the most important factors (PSFs) that need to be improved (e.g. increase situational awareness, improve human system interface quality, etc.) to avoid said human error. The results effectively are also indicative of the reality (the accident investigation report) and expectations of experts.

To address the issue of lack of data in HRA modeling multiple sources (empirical data, and expert data), and different types of data (objective, and subjective) are needed. Towards this objective, we aim to demonstrate the feasibility of PRELUDE extension with empirical data from simulator sessions. The second part of this thesis proposes (1) a protocol to obtain empirical human reliability data from

simulator experimentation, (2) propose a transformation and data analysis methods to augment the PRELUDE methodology.

A European Railway Traffic Management System (ERTMS) operational simulator was employed for this task. The experimental set-up is a track section (created on the simulator) and tasks required from the train driver. A scenario run is defined where a train driver has to accomplish some fixed objectives, in certain conditions associated to PSFs states. Since, HRA is mainly concerned with certain degraded conditions which can lead to a higher probability of human error (distraction, bad communication, etc.), thus, multiple scenario runs are defined, where each run aims to simulate a PSF in a degraded state. A selected number of PSFs that are important for operational safety in rail domain [Rangra et al., 2015a] are used. PSF's definitions, and inspiration from real world cases are taken into account to simulate the degraded conditions. Subjects are then invited to complete the simulation runs. For each run, objective human performance data is saved from the simulator. After each run, subjective self-assessment data is also obtained using standardized multi-scale and simple questionnaires. To analyze the objective data, criteria which links human performance, and the system level goals were chosen. This analysis aims to identify the effect of a PSFs' state (as created in the scenario) on human performance. In the first analysis, the objective data: score is transformed and combined with expert data to update the HRA model in VBS. Subsequently, subjective data results are presented and analyzed. The subjective data verifies whether subjects indeed perceived a degraded state of PSFs. Finally subjective and objective data are analyzed to identify other PSFs. This work also proposes to identify PSFs – self-estimation – from experimental data. Such, factors which were not analyzed or identified from a pure safety perspective but can occur under certain conditions, and can degrade human performance.

A list of publications during this thesis work are given as follows:

- Rangra, S., Sallak, M., Schön, W. and Vanderhaegen, F., 2017. A Graphical Model Based on Performance Shaping Factors for Assessing Human Reliability. *IEEE Transactions on Reliability*, 66(4), pp.1120-1143.
- Rangra, S., Sallak, M., Schön, W., & Vanderhaegen, F. (2015). On the study of human reliability in transportation systems of systems. *In 2015 10th System of Systems Engineering Conference (SoSE)* (pp. 208–213). San Antonio, TX, USA: IEEE.
- Rangra, S., Sallak, M., Schön, W., & Vanderhaegen, F. (2015). Human Reliability Assessment under Uncertainty – Towards a Formal Method. *In*

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015 (Vol. 3, pp. 3230–3237). Elsevier B.V. Procedia Manufacturing.

- Rangra, S., Bader, K., Sallak, M., Schön, W., & Vanderhaegen, F. (2016). Analyse de la fiabilité humaine: vers un cadre plus formel pour les applications ferroviaires. *In 20ème Colloque National de Maîtrise des Risques et Sûreté de Fonctionnement*, Lambda Mu 20, Oct. 2016. Saint Malo: Lambda Mu20.
- Rangra, S., Schön, W., & Vanderhaegen, F. (2016). Integration of human factors in safety and risk analysis of railway operations: issues and methods from the perspective of a recent accident. *In International Railway Safety Council (IRSC 2016)*. Paris, France.
- Rangra, S., Sallak, M., Schön, W., & Vanderhaegen, F. (2017). Obtaining empirical data from experimentations on railway operational simulator for human reliability modelling. *In Safety and Reliability – Theory and Applications, ESREL 2017* (pp. 50–50). CRC Press Taylor & Francis Group CRC Press.
- Rangra, S., Bader, K., Sallak, M., & Schön, W. (2017). Railway incident analysis using event tree and operational simulators: application for ERTMS operational rules. Submitted to *Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*.

Rest of this manuscript is structured as follows:

Chapter 2 gives the theoretical background of the present work. It starts by presenting some basic notions of risk, and reliability analysis and some methods which allow performing the said analysis. The mathematical framework that is employed to represent and manage uncertainty in data, i.e. the BFT-VBS framework and related data combination rules conclude this chapter.

The Chapter 3 presents the background and application overview of the problem-set that this thesis aims to address. It presents the notion of human error and its role in the system-of-systems view of transportation systems. It then presents the technical and regulatory details of the application context of this work.

Chapter 4 presents a comparative, critical state-of-art of HRA, mainly quantitative methods to identify the recurring notions and good practices. Subsequently, this focuses on the quantitative aspects of an HRA model such as the use of probabilistic graphical models, uncertainty in quantification, etc. This discussion is followed by a focus on rail applications existing methods in the research community, regulatory, and industry are presented and discussed to identify the challenges that remain to be addressed.

Chapter 5 proposes an original complete HRA methodology titled PRELUDE, its underlying framework, and its application on the case study. It is a quantitative and qualitative HRA methodology, applied to railway operations.

Chapter 6 presents the feasibility study of PRELUDE's extension with data from simulator experimentation. It presents a protocol to obtain empirical human reliability data from simulator experimentation. The simulator sessions with subjects are also presented, followed by objective and subjective data, and analysis results.

Chapter 7 concludes this thesis manuscript with some general conclusions, and perspectives for future work.

Theoretical background

Contents

2.1 Reliability and risk analysis basic notions	14
2.1.1 Some reliability and risk assessment methods	15
2.1.1.1 Fault trees	16
2.1.1.2 Failure modes, effects and criticality analysis	16
2.1.1.3 Event tree	16
2.1.1.4 Bayesian networks	17
2.1.1.5 Valuation-based systems	18
2.2 Mathematical framework to represent uncertainty	20
2.2.1 Belief Functions Theory	21
2.2.2 Some BFT-based combination rules	24
2.2.2.1 Dempster's rule	25
2.2.2.2 Yager's rule	25
2.2.3 Comparison between BPAs using interval and distance metrics	26
2.3 Conclusions	27

2.1 Reliability and risk analysis basic notions

Risk is traditionally defined as a combination of the probability or likelihood, and the consequence of a negative outcome of an event. An expected value of risk for that event can thus be calculated as the probability of occurrence multiplied by the consequences of its occurrence. For a safety-critical system some events (say a failure) can have a catastrophic consequences, and a high probability of occurrence. Both of these elements combined allow determining the risk. [Figure 2.1](#) shows an example of a classic risk matrix for determining the severity of a hazard for the system under scrutiny. Thus, to reduce the expected risk: either to reduce the probability or consequences or both, i.e. a *high* frequency and *catastrophic* event has a intolerable risk level.

Dependability is property that allows users to have a justified confidence in the service delivered by a system [[Laprie, 1992](#)]. There are various attributes of a system used to represent its dependability. These can be assessed to determine its overall dependability, although such notions date back to the 1980's [[Villemeur, 1988](#)], [[Laprie, 1992](#)], we cite a more recent work [[Avizienis et al., 2004](#)]. These attributes are given as follows:

- **Reliability:** continuity of correct service, under given conditions for a given time interval.
- **Availability:** readiness for correct service.
- **Maintainability:** ability to undergo modifications and repairs.
- **Safety:** absence of catastrophic consequences on the users and the environment.

*Frequency of occurrence of hazardous event	Risk Levels			
Frequent	Undesirable	Intolerable	Intolerable	Intolerable
Probable	Tolerable	Undesirable	Intolerable	Intolerable
Occasional	Tolerable	Undesirable	Undesirable	Intolerable
Remote	Negligible	Tolerable	Undesirable	Undesirable
Improbable	Negligible	Negligible	Tolerable	Tolerable
Incredible	Negligible	Negligible	Negligible	Negligible
	Insignificant	Marginal	Critical	Catastrophic
	Severity Levels of Hazard Consequence			

Figure 2.1 – A classic risk matrix for determining the severity of a hazard for the system under scrutiny [[CENELEC, 1999](#)]

-
- **Integrity**: absence of improper system alterations.
 - **Security** is defined as the combination of **availability** for authorized actions, **confidentiality** - the absence of unauthorized disclosure, and **integrity** the prevention of unauthorized alterations.

There are some threats to dependability: failures, errors, and faults. A **failure** occurs when the service delivered by the system deviates from what is required. The cause of a failure is an **error** affecting a part of the system's state. The cause of an error is a **fault**. The definitions are recursive because a failure of a component is a fault for the system containing that component. The causal chain is therefore: fault → error → failure

Several means are developed to attain the various attributes of dependability, as given below:

- **Fault prevention** means to prevent the occurrence of introduction of faults.
- **Fault tolerance** means to avoid service failures in the presence of faults.
- **Fault removal** means to reduce the number and severity of faults.
- **Fault forecasting** means to estimate the present number, the future incidence, and the likely consequences of faults.

RAMS is an acronym for Reliability, Availability, Maintainability, and Safety. RAMS or *reliability engineering* in general aims to develop methods and tools to evaluate and demonstrate dependability attributes. As [Høyland and Rausand, 1994] notes “*If safety and security are included in the definition of dependability as influencing factors, dependability will be identical to the RAMS concept*”. RAMS activities are generally integrated in the development life cycle of a product or service [Biolini, 2014]. Some of such tools are briefly described in the next section.

2.1.1 Some reliability and risk assessment methods

To perform reliability and risk assessment some models use graph-based representations, others use qualitative analysis of systems and components. The well known probabilistic graphical models offer a basis for representing compactly the probabilistic interactions between variables [Almond, 1995]. Their ability to reason using logic and probability and a graphical view offer ease of usage by non-experts, thus making them a good candidate for some applications.

2.1.1.1 Fault trees

A Fault Tree (FT) is a graphical representation of a system using Boolean logic. A fault tree analysis (FTA) aims to evaluate the state of a system, represented by the state of a top level event in terms of the states of basic events using Boolean logic. A top level event, which is generally a system failure is progressively decomposed into combinations of more simple events until a level where events are considered as elementary e.g. a failure of a component. FTA is a top-down (deductive) approach to reason about the system's safety. These elementary events are also known as basic events. It can be used for example to decompose a systems function into elementary functions and then provide the system's failure probability as a function of the failure probability of the elementary functions.

2.1.1.2 Failure modes, effects and criticality analysis

FMECA (or a simpler FMEA, without the criticality analysis) is a typical example of an inductive method, a bottom up approach. FMEA allows identification of the safety critical items which lead to severe consequences, but also latent failures (not immediately detected) which are good candidates to be part of multiple failures scenarios [Biolini, 2014]. The overall objective is to examine each potential component failure, and decide which components should be the focus of reliability improvement efforts in order to reduce risk as much as possible.

Thus, a FMECA is a methodology to identify and analyze:

- All **single failures and modes** (e.g. a component) and their consequences.
- The **effect** of these failures on the system
- A preliminary estimation of their occurrence **probability** to determine their **criticality**
- How to avoid the failures, and/or **mitigate the effects** of the failures on the system.

2.1.1.3 Event tree

In most safety-critical systems, a number of safety functions, or barriers, are provided to stop or mitigate the consequences of potential accidental events. The safety functions may comprise technical equipment, human interventions, emergency procedures, and combinations of these.

Event Trees Analysis (ETA) also called incident sequence analysis [Villemeur, 1992] is a bottom-up approach. An event tree is a logic tree diagram that starts from

a basic initiating event and provides a systematic coverage of the event propagation to its potential outcomes or consequences. It applies in particular for risk analysis of large systems with interacting internal and external factors [Høyland and Rausand, 1994].

An event tree analysis starts with an *initiating event* or potential accidental events and builds a tree of potential consequences depending on the subsequent events. The purpose of the method is to identify event sequences and their potential consequences. It particularly considers the effects of mitigation introduced to limit the effect (consequences) of the initiating event.

Each path from the root i.e. the initiating event to a leaf is an event sequence. Typically, a safety-critical system will have several layers of defense in order to control or limit any damage due to faults within the system. Hence, a sequence (or combination) of safety-related systems failure will typically constitute a critical event sequence in an event tree analysis.

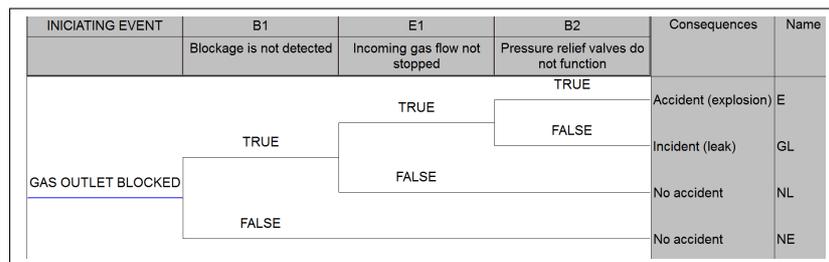


Figure 2.2 – An example of Event tree analysis: a gas pipe carrying a flammable gas

2.1.1.4 Bayesian networks

Probabilistic graphical models such as Bayesian networks (BNs) are also of particular interest in reliability analysis [Langseth and Portinale, 2007] [Weber et al., 2012], [Mkrtchyan et al., 2015]. Such models present some notable qualities when dealing with uncertainty and decision making [Aven and Zio, 2011] [Su et al., 2015].

In the BN interpretation, probability is considered as a belief about the occurrence of an event. Finally, the interpretation of the term probability signifies a degree of belief in the truth of a proposition, as determined from the data available. A BN is a probabilistic graphical model that represents a set of random variables and their conditional dependencies using a directed acyclic graph. It was developed as a framework for representing and evaluating models under uncertainty [Pearl, 2014].

The topology of a BN represents the variables that are conditionally independent given another variable. For example, in **Figure 2.3** X_2 is conditionally independent

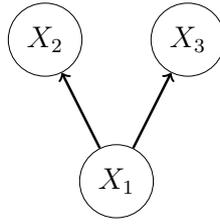


Figure 2.3 – BNs of a simple structure representing relation between three variables

of X_3 , given X_1 . The advantage of BNs is that they provide a compact representation of the joint probability distribution of the variables. This probability can be expressed as a product of the conditional distributions of each node given its parents in the graph.

Such a way to reason using logic and probability is called probabilistic reasoning. The input variables are instantiated and their probabilities are propagated through the network to update the probabilities of other nodes. The propagating procedure is based on Bayes' theorem and the structure of dependencies in a Bayesian network. One of the ways this reasoning is used is to represent a causal relationship: $X \rightarrow Y$, where X is a cause of Y and Y is an observable effect of X . The posterior probability distribution $P(X|Y = y)$ given the observation $Y = y$ can be computed using the prior distribution $P(X)$ and the conditional probability distribution $P(Y|X)$. The reasoning is performed using Bayes' rule, which is expressed in the following form:

$$P(X|Y = y) = \frac{P(Y = y|X)P(X)}{P(Y = y)} \quad (2.1)$$

where $P(Y = y) = \sum_x P(Y = y|X = x)P(X = x)$.

As described above, a BN contains two parts: the directed acyclic graph, and the quantitative part consisting of a joint probability distribution that factorizes into a set of conditional probability distributions governed by the structure of the directed acyclic graph.

2.1.1.5 Valuation-based systems

Valuation-Based Systems (VBS) was first defined in [Shenoy, 1989], and later in [Shenoy, 1992]. Similar to other probabilistic graphical methods VBS allow representing compactly the probabilistic interactions between variables [Almond, 1995]. They are not as popular as Bayesian methods, but do offer some notable qualities for the domain of risk and reliability analysis [Aguirre et al., 2013a] [Qiu et al., 2015] [Qiu et al., 2017].

This framework offers adequate tools to work with uncertainty in data and

experts. VBS can also be used to represent several domains for combination of the information such as: Bayesian probability theory, possibility theory, BFT, propositional logic, etc. This graphical view offers an easy visualization and usage by non-experts in BFT, and an intuitive display of results. Since, VBS can represent knowledge in different domains; it is possible to express valuations using basic probability assignments (BPAs), possibilities, and so on. Present work uses BPAs as presented in the section [section 2.2.1](#). More details on how BFT and VBS notions relate can be found in [[Shenoy, 1994](#)].

A simple example is used here to describe proposed interpretations of the variables and valuations (direct and configuration belief structures, [section 2.2.1](#)) in VBS. The variables are called HFE and PSF , the *variables* themselves are not introduced here, for now these are simple variables. Further, the relation between these variables is defined using *valuations*.

In a VBS's graphical representation, variables are represented by elliptical nodes, and valuations are represented by diamond-shaped nodes, as shown in [Figure 2.4](#). Here the set of variables of interest are $E = \{HFE, PSF_1, PSF_2\}$, where an HFE is the variable of interest, PSF_1 and PSF_2 are the other variables. Their respective frames are defined as $\Omega_{HFE} = \{true, false\}$, and for each of the PSFs as $\Omega_{PSF_i} = \{nominal, poor\}$. The frames are comprised of finite discrete values the variable can take. The relation between the PSFs and HFE is defined by using a *configuration belief structure*, the *BPA* represented graphically as m_1 in [Figure 2.4](#). It is defined on the frame $\Omega_\Phi = \Omega_{HFE} \times \Omega_{PSF_1} \times \Omega_{PSF_2}$. The other BPAs m_2 and m_3 contain evidence on the variables PSF_1 and PSF_2 respectively. As discussed before, they are *direct belief structures* and are used to represent data on single variables. These direct and configuration belief structures are then first combined, and then marginalized on the variable of interest (HFE) to obtain the quantification results. Under BFT-based usage of VBS, any one of various combination methods can be employed (as presented in [section 2.2.2](#)).

The quantification results i.e. upper and lower bounds are obtained by a combination of all the *BPA* and a projection on Ω_{HFE} .

This concludes the brief explanations of some reliability analysis methods: with underlying logical and or probabilistic reasoning. However, in some cases where there is a lack of data both in terms of the logical relations and probability, some of these methods cannot be used, and particular considerations need to be made for uncertainty. The work in this thesis deals with such problems where there is a lack of data, thus some specific mathematical frameworks are needed to manage uncertainty. Some such specific frameworks, are discussed in the following section.

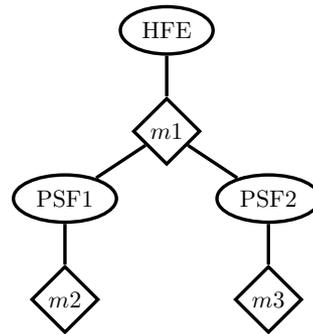


Figure 2.4 – A small example and the associated variables modeled as a VBS (a VN)

2.2 Mathematical framework to represent uncertainty

As the authors in [Aven and Zio, 2011] state: “the purely probability-based approaches to risk and uncertainty analysis can be challenged under the common conditions of limited or poor knowledge on the high-consequence risk problem, for which the information available does not provide a strong basis for as specific probability assignment: in such a decision making context, many stakeholders may not be satisfied with a probability assessment based on subjective judgments made by a group of analysts. In this view, a broader risk description is sought where all the uncertainties are laid out plain and flat with no additional information inserted in the analytic evaluation in the form of assumptions and hypotheses which cannot be proven right or wrong. This concern has sparked a number of investigations in the field of uncertainty representation and analysis, which have led to the developments of frameworks [i.e. alternative approaches for representing and describing uncertainties]”.

The context of this thesis also deals with some of these issues, hence this section introduces one such alternative approach to represent and manage uncertainty in data, which will be used later in this work. Note that there are other representations (probability bound analysis, imprecise probability, fuzzy probabilities, etc.) [Aven, 2011] we only describe the ones that we will use in present work.

For the estimation of probabilities of occurrence of some events owing to a lack of data on their failure rates some special considerations are needed to be made [Aven, 2011], in particular, the representation and management of uncertainty. We start by classifying the uncertainty in two types. This classification also popular in the domain of risk analysis [Aven, 2011], is also stressed in PRA (Probabilistic Risk Analysis) of complex systems [Parry, 1996].

Aleatory uncertainty (or variability, or stochastic uncertainty or irreducible uncertainty) is the physical variability present in the system being analyzed or its

environment – the natural variability of a variable. The determination or prediction of physical and operational condition of a physical system is typically of this type. Additional experimentation and/or characterization might provide more conclusive description of the variability but cannot eliminate it completely. Unless the environment is severely restricted, the event is isolated; this variability cannot be completely eliminated, therefore the term irreducible (uncertainty). The second type, epistemic uncertainty is defined as a *lack of knowledge about the 'true' value of the chances and parameters of the probability models* [Parry, 1996] [Aven, 2011]. It therefore represents questions not on the variable itself but the way a value of the variable is predicted.

2.2.1 Belief Functions Theory

Dempster-Shafer theory also known as evidence theory or Belief Functions Theory (BFT) was first proposed in [Dempster, 1967] and later extended in [Shafer, 1976]. It allows usage of upper and lower probability instead of precise values, with generalization of the Bayesian theory of subjective probabilities. The main contents of this theory are the combination and representation of evidence or knowledge. Evidence can be represented by a basic probability (belief) assignment and distinct pieces of evidence are combined by using a combination rule. It can represent and propagate both (epistemic and aleatory) types of uncertainty. It has been applied to different domains of application: data fusion [Smets, 1999], reliability and risk analysis [Sallak et al., 2013] [Qiu et al., 2015], and some aspects of HRA [Su et al., 2015]. A model based on BFT can be represented with an Evidential Network (EN) – a probabilistic graphical model. The basic elements of BFT framework are briefly described below:

Variables and configurations: A finite set of variables is used to model the problem at hand. Let's represent this set of all the variables in the problem by $E = \{X_1, X_2, \dots, X_n\}$. For each decision making problem, inference is then drawn only on a reduced domain of interest Φ . For a variable X , its frame Ω_X holds all possible values of this variable. Further, for a finite non-empty sub-set of variables $\Phi \subseteq E$, Ω_Φ denotes the Cartesian product of Ω_{X_i} for X_i in Φ : $\Omega_\Phi = \times\{\Omega_X | X \in \Phi\}$. Here Ω_Φ is called the frame (of discernment) for Φ . The elements of Ω_Φ are considered as configuration of Φ . For example, a set of variables $\Phi = \{X_1, X_2\}$, and their respective frames are: $\Omega_{X_1} = \{a_1, b_1\}$ and $\Omega_{X_2} = \{a_2, b_2\}$, then the frame of discernment for the configuration becomes $\Omega_\Phi = \{(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2)\}$.

Valuations and Basic Probability Assignments (BPA): A valuation m^{Ω_Φ} holds the knowledge about the possible values of variables in Φ . The set of valuations is denoted by $\Psi_\Phi = m^{\Omega_\Phi} : \Phi \subseteq E$. A valuation represented by m^{Ω_Φ} is used to represent knowledge about the possible values (or sets thereof) of Φ . The set of all the valuations in the problem set Ψ_E is further divided into two types of valuations: the direct valuations (posterior or input) Ψ_D holding the evidence about the input variables; and the prior domain Ψ_P holding the valuations that relate the variables amongst themselves. In this work direct valuations are valuations on singletons of E and prior assignments contain at least two elements of E .

In belief functions theory, valuations correspond to either basic probability assignment functions, belief functions, plausibility functions, or commonality functions. For simplicity, we describe belief functions theory in terms of *basic probability assignment*.

Basic Probability Assignments also known as a *mass function* is a way of representing confidence in a certain proposition. That is, the confidence that X is equal to a certain numerical or a linguistic value in its frame of discernment. For a variable of interest X and its frame, Ω_X , the mapping $m^{\Omega_X} : 2^{\Omega_X} \rightarrow [0, 1]$ that assigns values to the non-empty sets of the power set 2^{Ω_X} in the interval $[0, 1]$, is called a basic probability assignment. Further, these mappings are such that for a set A of the power set $2^{\Omega_X} : \sum_{A \subseteq \Omega_X} m^{\Omega_X}(A) = 1$, $m^{\Omega_X}(A) \geq 0$ and $m^{\Omega_X}(\emptyset) = 0$. Here, A , is a subset of Ω_X with nonzero values of m is called a focal set. A BPA: $m^{\Omega_X}(A)$ hence, reflects the degree of belief (subjective probability) committed to that part of the evidence which exactly points to A and A only. To note that A can either be a singleton (a single value $A = \{a\}$) or set of elements ($A = \{a, b\}, \{a, b, c, \dots\}$). A given BPA is similar to a probability function if the focal sets are singletons ($m(A) : A = \{a\}$). Further, BPA: m^Ω is assigned to each subset of 2^Ω instead of Ω , same as in classical probability theory. Therefore, each focal set has a BPA (strictly positive) based on the evidence about that focal set. Complete ignorance about X , that is, absolutely no knowledge about the true value of X , is represented as the BPA assigned to the whole frame i.e. Ω_X . It is represented as $m^{\Omega_X}(\Omega_X)$ or simply belief assigned to Ω_X . This also extends towards defining relation between two variables (a configuration). For example, let X and Y be two variables with frames $\Omega_X = \{a, b\}$ and $\Omega_Y = \{c, d\}$ respectively.

The relation between X and Y is represent as a joint belief or joint valuation, defined on frame $\Omega_{XY} = \Omega_Y \times \Omega_X$ that is $\Omega_{XY} = \{(a, c), (a, d), (b, c), (b, d)\}$ [Xu and Smets, 1996]. Here, the joint BPA assignment $m^{\Omega_{XY}}$ is used to represent the belief about the possible relation(s) given by a subset of $2^{\Omega_{XY}}$, between variables X and Y . For example, $m^{\Omega_{XY}}(\{(a, c)\}) = 0.8$; $m^{\Omega_{XY}}(\{(b, c), (b, d)\}) = 0.2$, are some BPAs

representing joint valuations, given that they respect the conditions of its definition.

In [Shafer, 1976], there is a constraint that a mass function must not assign a positive value to the empty set $m(\emptyset) = 0$. A mass function satisfying this property is called normalized. In [Smets, 1992] the authors propose that $m(\emptyset) = 0$ corresponds to a closed-world assumption which means the true value is included in the frame of discernment, while $m(\emptyset) > 0$ corresponds to an open-world assumption which means that the true value may not included in the frame of discernment. This work considers only closed-world assumptions. This means the frame of discernment is assumed to be exhaustive. Thus, using the normalization operation, un-normalized mass functions can be transformed into normalized mass functions as follows:

$$m'(A) = \begin{cases} \frac{m(A)}{1 - m(\emptyset)} & \text{if } A \neq \emptyset \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

In this work, the term *direct belief structure* will be used when talking about evidence on single variable X , for its frame Ω_X . And *configuration belief structure* for referring to joint valuations which represent relational evidence between two or more variables. That is, for X and Y , the BPA $m^{\Omega_{xy}}$ and corresponding focal sets will be the constituents of a *configuration belief structure*. This structure formally defines the relation between X and Y . Hereafter, these terms *focal set* and *assigned belief or BPA value* are used to refer to the contents of a *belief structure*.

Upper and lower probability bounds: After combination a marginalization is a projection on the frame of the variable of interest (i.e. it aims to crystallize the available combined knowledge on elements of Ω_X for the variable of interest X). Intuitively, combination corresponds to aggregation of knowledge and marginalization corresponds to crystallization thereof [Shenoy, 1989]. These marginalized results he results obtained can be interpreted in the form of a lower (P_{inf}) and upper bound (P_{sup}) or measure on the variable's values, as per Dempster's interpretation [Dempster, 1967]. In other interpretations this upper bound corresponds to a plausibility function in belief functions and the lower bound to belief function, this work to keep interpretations simple employs the notions of upper and lower bound. For two subsets A and B of the variable of interest X , they are defined as below:

$$\begin{aligned} P_{inf}(A) &= \sum_{B \subset A} m(B) \\ P_{sup}(A) &= \sum_{B \cap A \neq \emptyset} m(B) \end{aligned} \quad (2.3)$$

Here the length of the interval (i.e. $P_{sup}(A) - P_{inf}(A)$) or belief on the frame

$m(\Omega_X)$ represents this imprecision (epistemic uncertainty) about A . Finally, it can be said that actual probability that the value of X belongs to A (any of the elements of A if A is not a singleton) is included in the closed interval composed of the lower and upper bounds. Furthermore, $P_{sup} > P_{inf}$, and if there is an absence of epistemic uncertainty then $P_{sup}(A) = P_{inf}(A) = P(A)$.

To note that a Bayesian belief structure corresponds to the classical probability model in which the focal element is singletons, $A = x$. In this case $m(A)$ corresponds to the probability that x , is the value of A . Furthermore, note that all these formulas are compatible with the Bayes rule.

2.2.2 Some BFT-based combination rules

Once the knowledge is represented it has to be *combined* to make inferences. Combination rules allow aggregating all the mass functions into a combined mass function. We work on valuations in [section 2.2.1](#) to represent information, thus they are the objects to be combined or marginalized. A mapping $\oplus : \Psi_1 \times \Psi_2 \rightarrow \Psi$ is called combination which aims to aggregate knowledge. The combination of multiple valuations: $\oplus\Psi$, is called the combined valuation. And as discussed before the masses to be combined should be defined on the same frame of discernment. However, if that is not the case the mass on distinct variables X and Y must therefore be extended to the product space of $X \times Y$ by an operation called vacuous extension. The work in [[Smets, 1993](#)] introduces the principle of minimal commitment, which allows the construction of new belief functions on refined spaces (a vacuous extension). [Equation 2.4](#) shows the extension of a bpa m^{Ω_X} to Ω_{XY} , by transferring each bpa $m^{\Omega_X}(B)$ to the extension of $B : B \times \Omega_Y$

$$m^{\Omega_X \uparrow \Omega_{XY}}(A) = \begin{cases} m^{\Omega_X}(B), & \text{if } A = B \times \Omega_Y. \\ 0, & \text{otherwise.} \end{cases} \quad (2.4)$$

Within the framework of BFT there are several combination rules that allow combination of knowledge held by several pieces of evidences. Intuitively, combination corresponds to aggregation of knowledge and marginalization corresponds to crystallization thereof [[Shenoy, 1989](#)].

Since in this is also interested in BFT from the perspective of combination of data, thus considerations on the nature of data or information from experts are also discussed here. It may be noted that, these combination rules can be used irrespective of the source of data (expert, empirical, etc.) as long as their respective assumptions are fulfilled. Suffice to say, there exists plenty of other combination rules such as minimal commitment principle, conjunctive, disjunctive, etc., offering

an interpretation and modeling of the knowledge. A comprehensive discussion can be found in [Sentz and Ferson, 2002]. A select few that are used in the present work are described below.

2.2.2.1 Dempster's rule

Dempster's rule of combination (also known as product-intersection rule) was introduced in [Dempster, 1967] and interpreted later in [Shafer, 1976]. It combines normalized mass functions over the same frame of discernment.

Let m_1 and m_2 be normalized mass functions on the same frame of discernment Ω . The combined mass function $m_{1\oplus 2} = m_1 \oplus m_2$ is defined as:

$$m_{1\oplus 2}^\Omega(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-k} \sum_{\phi \neq B \cap C = A} m_1(B)m_2(C) \quad (2.5)$$

where: $\forall A, B, C \subseteq \Omega, A \neq \{\phi\}$, and $k = \sum_{B \cap C = \phi} m_1(B)m_2(C)$ is a measure of the conflict between the two *bpa*, i.e. m_1 and m_2 . The usage of $1-k$ (also known as normalization factor) is such that it takes this conflict into account by redistributing or normalizing this value. This is performed by redistributing the mass assigned to the empty set $m(\phi)$ uniformly amongst all the other masses except the empty set itself. If $k = 1$, it represents a complete conflict and Dempster's rule is not defined for this case. Thus, if completely opposed masses are there Dempster's cannot be used. If $k = 0$ the sources are completely in agreement. Further, Dempster's rule is commutative, associative, and not idempotent.

In this rule, it is assumed that all masses stem from fully reliable and independent sources. It has widespread usage partially because of its ease of application. It considers that the data sources are equally reliable and independent. Furthermore, it manages small conflicts by redistributing the conflicting *BPA* in a uniform way to other focal elements using a normalization factor $1 - k$ (where k is a measure of the degree of conflict). Note that other rules are defined when the sources are not independent or reliable.

2.2.2.2 Yager's rule

Yager's rule [Yager, 1987] was introduced as a modification of Dempster's rule to address among others, the normalization factor leading to counter intuitive results in cases of highly conflicting evidence. It has two main differences to Dempster's rule: firstly the author argues that an important feature of combination rules is the ability to update an already combined structure when new information becomes available, it is interpreted as that a combination rule should be non-associative,

thus Yager's rule is non-associative or "quasi-associative" as stated in [Yager, 1987]. Secondly, it detects conflict similar to the Dempster's rule, however, it is managed by redistributing the conflicting BPA (K) to the frame of discernment ($m(\Omega)$) i.e., it considers conflict as an additional source of uncertainty). Such a management of conflict is justified by [Yager, 1987] as follows: "In this case we are saying that since we don't really know anything about the conflicted portion, we let it be distributed among all the elements rather than just those in the focal sets."

Let m_1 and m_2 be normalized mass functions on the same frame of discernment Ω . The Yager's combination operator is denoted here by \odot , further the combined mass functions is defined as $m_{1\odot 2}$ is given as:

$$\begin{aligned} m_{1\odot 2}(A) &= (m_1 \odot m_2)(A) = \sum_{\phi \neq B \cap C = A} m_1(B)m_2(C) \\ (m_1 \odot m_2)(\Omega) &= \sum_{\phi \neq B \cap C = \Omega} m_1(\Omega) \times m_2(\Omega) + k \end{aligned} \quad (2.6)$$

where: $\forall A, B, C \subseteq \Omega, A \neq \{\phi\}$, and $k = \sum_{B \cap C = \phi} m_1(B)m_2(C)$, similar to Dempster's rule. Note that in the case where $k = 0$, i.e. there is no conflict this rule gives same results as the Dempster's rule. Similar to Dempster's rule this rule also assumes that all the sources of information are independent and reliable. It is quasi-associative, commutative and non-idempotent. It may be noted that, when dealing with such frameworks reliable and independence assumptions are common [Podofilini and Dang, 2013].

2.2.3 Comparison between BPAs using interval and distance metrics

In the context of belief functions theory comparing given BPAs using a metric such as distance can be useful. This comparison can be used to measure similarity or dissimilarity between the information represented by two BPAs for applications such as clustering, classification, etc. [Jousselme et al., 2001]. Such a pairwise comparison can be interesting when BPAs are obtained from different information sources (sensors, experts, etc.) or after using different treatments (such as after combination rules). This distance metric complements the usage of BFT by giving the user a tool to interpret the degree of (non-)alikehood between belief functions in a meaningful way [Loudahi et al., 2014]. There are different approaches to compare two BPAs, interested readers can refer to works such as [Loudahi et al., 2014] [Cuzzolin, 2008].

In the present work's usage of BFT another rather straightforward approach is

to use *the middle of the interval* for a given upper and lower probability bound (also known as the pignistic probability). Such a metric is often used at the decision-making-level in the context of belief functions. In simpler usage it can also be used to compare BPAs, such as the ones obtained after using different combination rules [Sebbak et al., 2014]. If different BPAs for a same frame are to be compared, the middle of the interval for a value of interest (e.g. a HFE being true) can be used. It can be limiting in the sense that only a variable's value of interest can be compared across different BPAs.

Secondly, in a more general sense, two BPAs on the same frame can be compared using distance metrics. One of the rather well-known distance measures is *Jousselme distance* [Jousselme et al., 2001]. It proposes the use of a classical similarity measure to achieve the comparison of two BPAs. Let $m1$ and $m2$ be two BPAs on the same frame Ω . Then the Jousselme distance d_J between $m1$ and $m2$ is defined as follows:

$$d_J(m1, m2) = \sqrt{\frac{1}{2}(\|m1\|^2 + \|m2\|^2 - 2\langle m1, m2 \rangle)} \quad (2.7)$$

where $\langle m1, m2 \rangle$ is the scalar product defined by:

$$\langle m1, m2 \rangle = \sum_{i=1}^n \sum_{j=1}^n m1(A_i) m2(A_j) \frac{|A_i \cap A_j|}{|A_i \cup A_j|} \quad (2.8)$$

where $n = |2^\Omega|$ and A_i and A_j focal sets of all the pieces of information represented by $m1$ and $m2$ respectively, and $\|m1\|^2$ the square norm of $m1$.

As a simpler form of comparing different BPAs for a given variable's value of interest, this paper uses the middle of the intervals. Secondly, the distance metrics presented here d_J will also be computed to compare two complete sets of information represented by two given BPAs.

2.3 Conclusions

This chapter has introduced some basic notions of risk assessment. Some methods used to perform these assessments were also briefly discussed. These definitions and notions have been used over the years to assure dependable systems and their safe operations. Frameworks such as FTA, Event tree, etc. have seen industrial usage towards ensuring the various RAMS attributes. Their ability to reason using logic and probability and a graphical view offer an easy visualization and usage making them a good candidate for various applications. However, the lack of data problem and traditional methods limited expressiveness, calls for special methods to model complex systems and represent uncertainty. For such problems frameworks like

Bayesian networks, BFT-VBS, etc., can be employed. The application context of this thesis are such systems or rather systems-of-systems, where on one hand the application and usage of traditional approaches is not straightforward; in addition to the lack of empirical data required to model the problem. The next section presents details of the problem background of this thesis.

Problem background

Contents

3.1 Labex MS2T: control of technological Systems-of-Systems . . .	30
3.2 Transportation system-of-systems and safety	31
3.3 Human errors: some notions	34
3.3.1 Some perspectives on analyzing human errors	34
3.3.2 Human error: classification and taxonomies	36
3.3.2.1 Person approach	37
3.3.2.2 Systems approach	38
3.4 Application context: railway operations	38
3.4.1 European Rail Traffic Management System	39
3.4.1.1 ERTMS application levels	40
3.4.1.2 ERTMS/ETCS braking curves and train driver DMI	41
3.4.2 European regulations and human errors	45
3.5 Conclusions	47

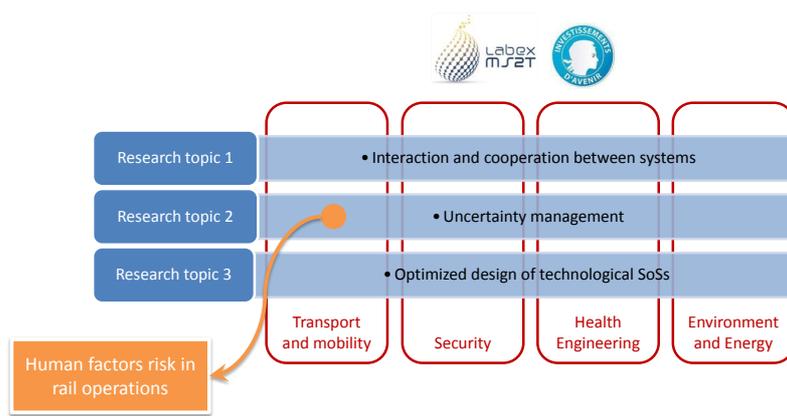


Figure 3.1 – An overview of *Labex MS2T* and the positioning of this thesis.

3.1 Labex MS2T: control of technological Systems-of-Systems

The *Labex MS2T* is a “*Laboratories of Excellence*” program, coordinated by ANR (French national research agency) and carried out at the UTC. The *Labex MS2T “Control of Technological Systems-of-Systems”*¹ is a multi-disciplinary problem that targets a large scale scope of application.

The unprecedented development of means of communication today requires a large-scale interconnection of autonomous technological systems that can cooperate to perform certain tasks. “Technological systems of systems” (TSoS), which are defined in particular by the autonomy and heterogeneity of the component systems. The Management of Technological Systems of Systems (MTSoS), which is the focus of this project, targets a potentially very wide scope of application, including very important socio-economic issues in the fields of :

- Transport and mobility
- Security
- Health engineering
- Environment and energy

The context of this thesis figures on the axis of two research topics – **transportation domain** and **uncertainty management**.

The next section details the problems of transportation system-of-systems, where particular focus is put on safety aspects.

¹<https://www.labexms2t.fr/>

3.2 Transportation system-of-systems and safety

One definition of SoS from the work in [Jamshidi, 2011] states: “a SoS is an integration of a finite number of constituent systems which are independent and operable, and which are networked together for a period of time to achieve a certain higher goal”. The Strategic Research Agenda (SRA) for research in Systems of Systems Engineering in the European Union (EU) [Henson et al., 2013] defines it as: “networking individual systems together to realize a higher goal that none of the individual systems can achieve in isolation” and furthermore also defines emergence as a fundamental property of SoS, a sentiment reciprocated over the years in other works. In [Wilson, 2014], the authors argue “. . . in essence the railway is a large, complex distributed system of many technical, organizational, economic and human components.” They establish some parallels between the SoS notions and that of human factors, for the railway domain namely: context, acknowledgment of interactions and complexity, a holistic approach, recognition of emergence and embedding of the professional effort involved within organization system. Thus, rail transport presents various traits of a SoS. However most of these notions are abstract in form. Rest of this section elaborates the context of thesis namely: the human and safety.

In [Barot et al., 2013] the authors list some “high level SoS problem areas.” They state that SoS problems are *wicked problems*, that is problems that are “extremely complex and not bounded or stable; they do not have unique, right solutions, but rather solutions that are either better or worse than others, and they do not have a definitive formulation; SoS requirements are often volatile with changing constraints and moving targets; stakeholders have different views; and understanding the whole context is difficult.” They further go on to list **Safety**, Security and Integrity as one of these high level problems for SoS. Human aspects figure predominantly when discussing the socio-technical aspects of SoS. Human considered as a component of a transportation System of Systems for risk assessment allows us to study its impact on system reliability and give feedback to improve overall safety [Rangra et al., 2015b].

Safety in a socio-technical system is sometimes seen as a control problem, and there are various actors involved to ensure the system remains safe. In [Rasmussen, 1997] an overview of the whole chain of different components of a socio-technical systems involved in the control of safety. Such a view shows the overwhelming picture of the control problem to ensure safety. The question then changes from *how to analyze safety*, to, *if it is possible at all to ensure safety at this level of complexity*?. A SoS view provides with adequate directions to handle this problem. Human

- studied as part of a socio-technical system for example a train or a car driver, a signaller etc. does exhibit most properties expected from the constituents of a System of Systems [Jamshidi, 2011]. When we consider human controller as a component of the system it exhibits autonomy, operational independence and induces emergent properties, co-operating with other components towards a common goal [Wilson, 2014]. Therefore, for most cases, interacting systems involving a human users can be treated as a SoS. Furthermore, railway signaling systems, ERTMS in particular have been shown to possess the properties expected from SoS in [Qiu et al., 2014]. A road vehicle with embedded control systems (constituents of an ADAS) also exhibits properties of a SoS [Samad and Parisini, 2011]. The second, prospective point of view also agree with this: accident analysis reports sometimes attribute an unforeseen interaction of subsystems (involving a human) as the cause of system failure. This relates to the emergent properties exhibited by a SoS.

Emergence remains a disputable topic in the SoS domain. There is a lack of precise and/or universally accepted definition. It is often seen from the eyes of an application domain, and various works define or interpret it as applicable to a problem-set. The work in [Jamshidi, 2011] considered emergence to be *“Something unexpected in the collective behavior of an entity within its environment, not attributable to any subset of its parts, that is present in a given view and not present in any other view.”* The source of this unexpected behavior is stated to arise from interaction between the components of an application and their environment [Johnson, 2006]. They further point out that emergent properties can be beneficial or they can be harmful if they reduce the safety requirements. Thus, an emergent property is seen as a **higher-level property** which stems from the interaction of **lower-level entities** and the **environment**, it cannot be deduced directly from the properties of lower-level entities. Discussing the challenges for the domain of reliability and safety [Zio, 2009] states: *“insights from research on failures in complex systems have revealed that safety is an emergent property of a system and their constitutive elements, rather than a resultant one.”* Thus, the higher-level property we focus on is **system safety**; the entity and their interaction is a human in interaction with other entities in assuring the service (the operational context) and a standard work environment is considered.

For the providers of large scale and complex services such as air or rail transport risk is inherent, it cannot be completely eliminated [Perrow, 1999] [Amalberti, 2001]. The more complex the systems become, the larger the scope and analysis of risk becomes [Rasmussen, 1997]. Likewise, owing to the increase in the number of interaction between components, analyzing the reliability of the sub-systems including a human, becomes difficult.

In traditional risk analysis of systems, the tools and methodologies are available to address defined problems. Since the system boundaries are fixed and expected behavior is known scoping these problems and the associated risks are relatively well-understood. However, for system-of systems where in some cases the definitions are not completely straightforward, risk management is a critical but immature element [Kinder et al., 2015]. Thus, for considering a SoS and associated risk, the boundaries of the analysis vs. traditional risk analysis are needed to be extended [Zio and Ferrario, 2013]. Similarly, for risk analysis in safety critical transportation SoS also needs similar extension, inwards as well as outwards. Quantified risk analysis methods have been said to be preferable option for SoS risk assessment [Kinder et al., 2015].

It has been widely reported in rail and road transportation that majority of accidents are caused at least in part, by some form of human error. More recently a study [Evans, 2011] concluded the broad causes: Signal passed at danger, over-speeding, signaling or dispatching error, i.e. primarily human functions accounted for around 70% of the accidents. Further, train drivers are said to contribute to approximately 75% of the accidents analyzed 1945-2012 [Kyriakidis et al., 2015b]. On the other hand an Australian study of over 100 rail accident reports found that accidents attributed to human errors were caused by the conditions in which drivers had to work, indicating driver errors were in fact consequences, not the initial issue causing the accident [Edkins and Pollock, 1997]. It has been well accepted that statements like – human error caused the accident – is an oversimplification at the very least [Sheridan, 2008]. The absence of certification requirements also affect the design of the system as human considerations are not imposed and the system is conceived with an independent design perspective, which further adds to overall risk during system operation by a human [Di Grazia et al., 2014].

Furthermore, in systems like railway, human operators are technically skilled professionals often with multiple years of experience, significant training and are regularly evaluated for job fitness. In such a context the problem of a human error becomes larger than individual issues [Sheridan, 2008]. It can become a question of systematic failure of the training, support, and evaluation measures in place. Clear deviations from the prescribed tasks, may be classified as human errors because they are not what is specified and asked of a human, however when looked at from a broader multi-criteria perspective, deviations and problem solving capabilities of a human controller does yield beneficial results.

To conclude, the following remarks can be made:

- Rail transport presents various traits of a SoS. And a human actor - exhibits most properties expected from the constituent of an SoS.

- Safety, security and integrity are one of the high level problems for SoS.
- Safety at the SoS-level is seen an emergent property; on the contrary accident scenarios are often described as unforeseen interaction of systems/subsystems – a negative emergent property.
- Humans are often said to be involved in such scenarios, however human (error) related risk analysis lacks concrete requirements for some domains.
- To analyze SoS and associated risk, the boundaries of the analysis vs. traditional risk analysis are needed to be extended.
- Quantitative risk analysis methods have been said to be preferable option for SoS risk assessment.

Thus, a particular focus on the underlying notions of human errors is needed to understand how to integrate them in the risk analysis process. The next section introduces some such notions.

3.3 Human errors: some notions

Human errors are held responsible for a large share of accidents causes across application domains. There are various domain specific accident studies: [Evans, 2011] [Gaur, 2005] [U.S. Nuclear Regulatory Commission, 2002] and most of them accept the fact that there is a high rate of human involvement in the accidents in some way or the other. However, most of them fail to agree on a single approach to mitigate this issue.

To connect risk and human error [Sheridan, 2008] propose a combination of various values: probability of the opportunity for an error, probability that the error is committed, and probability that no recovery is made before the undesirable consequence. Reducing the opportunity of an error and making recover possible often fit into the notions of traditional fault avoidance and fault recovery. Probability of committing the error is the focus of most quantitative human error analysis approaches. However, to understand how to quantify the probability of committing the error some notions on the SoS view are presented to understand the context in which these errors are to be considered.

3.3.1 Some perspectives on analyzing human errors

Traditionally speaking, the concept of human reliability confronts the problem of its definition. It can be defined as technical reliability, i.e. the ability of a (human)

component to realize its allocated functions successfully, in given operational conditions and over an interval of time. A measurement of this ability is usually the probability of success. However, this definition is not sufficient [Vanderhaegen, 2010]. The human reliability is not static but evolves dynamically regarding learning effects and cooperative activities [Vanderhaegen, 2011], and its assessment is rather multi-criteria than mono-criterion. It usually relates to tasks to be achieved by human operators instead of functions and to the characteristics of these tasks and of the human resources [Vanderhaegen, 1999b].

The human characteristics can be interpreted as constraints for achieving tasks. There are characteristics such as: overloaded or under-loaded or hypo-vigilant; experienced or inexperienced, etc. Some of these characteristics relate to the so-called Performance Shaping Factors (PSFs). These factors that may affect the system performance are numerous and correlations between factors has to be identified in order to simplify their integration into a human reliability assessment. Moreover, the main difference between humans and machines is the possibility that humans do not respect voluntarily a given prescription for specific reasons due to organizational factors for example, or to create new tasks or functions by using differently the technical resources [Vanderhaegen and Zieba, 2014]. In such cases, humans are not repaired or changed, but they adapt their own behaviors to specific or usual constraints they have to control. Thus, even if a human is considered as a functional component of a normal system, they do not necessarily adhere to the traditional notions of dependability (section 2.1).

On the other hand human reliability assessment can have several sources of explanation [Vanderhaegen, 2010]: the assessment made by the designers of a given human-machine system, by an industrial organization that will employ people in order to operate this system, and the assessment made by the users of such a system. Sometimes these assessments differ. The feedback of experience is then required in order to integrate the natural learning effects of human operators into the design process and to take into account the behaviors applied for controlling well-known or unprecedented situations. Joint prospective, retrospective and on-line approaches are useful in order to guarantee the efficiency of the human reliability assessment. Evidential networks or Bayesian networks can then be suitable tools to support such an assessment [Aguirre et al., 2013b],[Sedki et al., 2013].

One such example is dissonance engineering [Vanderhaegen, 2014b]. A cognitive dissonance is defined as an incoherence between cognitions. Cindynics dissonance is a collective or an organizational dissonance related to incoherenc between persons or between groups of people. Finally dissonance engineering is the treatment of such conditions. It is a concept which has applications in the

context of human errors, especially with the design of newer systems or integrating them with legacy systems, a case frequently seen in a domain like railway. Such cases give rise to contradictory and possibly safety critical conditions especially in operational conditions. The theoretical concept and an application example on car driving with the use of ADAS (Advanced driver assistance systems) is given in [Vanderhaegen, 2016].

When designing a SoS, a quantitative risk analysis identifies undesirable scenarios for which the designers have to specify material barriers or manual procedures in order to make them acceptable and reduce the residual risk level under a maximum value. This process does not consider that human operators can sometimes remove some of these barriers in order to optimize the compromise between performance criteria such as safety, task load, quality or production of service for instance [Polet et al., 2003] [Vanderhaegen et al., 2011]. The risk assessment of barrier removals is an interesting challenging topic but requires a strong collection of field data to develop relevant human behavioural models. For instance, models based on dissonance engineering can support the representation of rule or knowledge of a SoS functioning and use, and can identify possible dangerous or beneficial dissonances involving human, technical, environmental or organizational factors [Vanderhaegen, 2014a] [Vanderhaegen, 2016] [Vanderhaegen and Carsten, 2017] [Qiu et al., 2017].

Thus, there are various characteristics of human error analysis which cannot be defined or analyzed same as the traditional notions of RAMS attributes. The next subsections sheds some light on some of these characteristics.

3.3.2 Human error: classification and taxonomies

It has been widely accepted that modern accidents are not single cause events and generally tend to be sequences of undesired events (or decisions), bypassing multiple redundancy barriers and other safety features. One rather well known approach for complex socio-technical systems is the Swiss cheese model of accident causation [Figure 3.2](#). Such a view aims to visualize and assess the notions of barriers involved in an accident and identify holes in a complex socio-technical system's operation, such as railway operations.

In [Reason, 2000] two ways to consider human errors, are discussed: the person and system approaches. Each has its models of error causation and provides different theories of error management. The following discussion presents briefly these approaches.

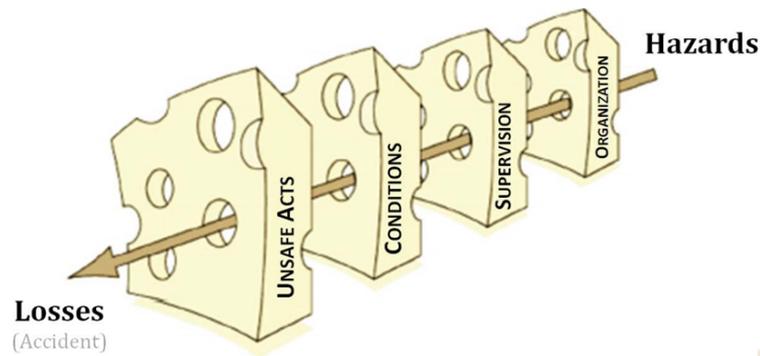


Figure 3.2 – A Swiss cheese model representing accident causation, from [Reason, 2000]

3.3.2.1 Person approach

The person approach focuses on the unsafe acts, as seen in [Figure 3.3](#) [Reason, 1990]. A focus tends to be on human behavior, error mechanisms with the objective of a systematic internal understanding of a human performance [Rasmussen, 1982]. Further, these approaches tend to have a focus on procedural violations of people, generally in domains like healthcare. With the view that these undesired acts primarily arise from the mental state of the human and subsequently call for a focus on human behavior [Reason, 2000]. Thus, the primary causes of such unsafe acts are said to be aberrant mental processes such as distraction, loss of concentration, memory lapses, poor motivation or decision making skills.

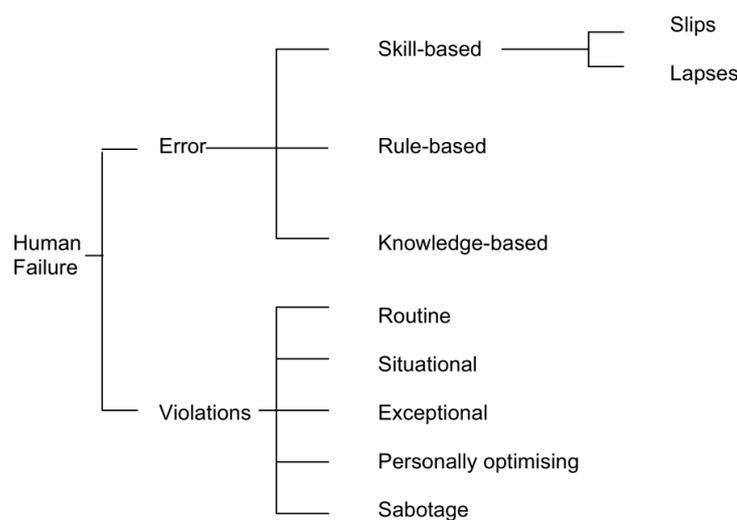


Figure 3.3 – Unsafe acts or human failures breakdown in a person approach to analyze a human error

3.3.2.2 Systems approach

The systems approach to human error states that “*humans are fallible and errors are to be expected, even in the best organizations*” [Reason, 2000]. Further, these errors are the inevitable consequence of inadequate conditions residing within complex systems. This approach is more recognized and used in a retrospective analysis, i.e. accident analysis, and forms the basis of so-called Systemic Accident Analysis (SAA) approaches [Underwood and Waterson, 2013] [Leveson, 2015]. In [Underwood and Waterson, 2014] the authors provide a comparison of such approaches by applying them to a railway accident. Notions like system safety as an emergent property, organizational, management influences are often the center point of these approaches.

The Swiss cheese model shown in Figure 3.2, is another type of systems approach since there is a consideration of all the levels as barriers, which under certain circumstances can lead to an accident [Reason, 2000]. Systemic models consider that an error or an accident are *emergent phenomena* which arises due to the complex interactions between system components that may lead to degradation of system performance, or result in an accident [Qureshi, 2007].

Systemic models, for a retrospective analysis provide a detailed and robust framework to investigate in detail accidents and human errors. The feedback provided with such models are more detailed and beneficial to investigative authorities. However, they can be complex to use, and require significant time, resources, and information [Underwood and Waterson, 2014].

The idea that for an SoS, safety and (on the contrary) accident (or errors) are essentially emergent phenomena is thus well accepted [Qiu et al., 2014], [Qureshi, 2007], [Leveson, 2011]. How to analyze and predict these safety aspects, specially for the human component forms the central premise of this thesis. Further to propose a more pragmatic approach, this works focuses on the lower levels of ?? (the work and staff-level), and operational conditions as seen in Figure 3.3. The next section details the application context in which we aim to apply our work.

3.4 Application context: railway operations

Rail transportation has multiple entities all contributing towards a safe and efficient transportation service. It is composed of multiple human actors (drivers, signalers, maintenance personnel, operational management) and signaling systems working in a synchronized way towards the achievement of some final goals. Railway signaling is one of the basic elements of railway operational safety [Schön et al.,

2013]. Human factors have always been an important considerations in railways, the guide [[Rail Safety & Standards Board, 2008](#)] presents a background on the transverse nature and complexity of HF considerations.

The context of application for this discussion is limited to *railway operations*, it is defined as: *a train movement from one point to another*. For our considerations of railway operations, we focus on the train driver as the subject, to limit our problem-set. Thus, this section describes the underlying signaling system, with a particular focus on the railway signaling–train driver information interface. Appropriate data and additional information is provided here which will be later used to analyze the human/train driver performance in its operational context. There are various signaling systems and standards in place in different countries all over the world. ERTMS (European Rail Traffic Management System) is a relatively new entrant, although fast gaining ground in Europe and elsewhere. Working within the framework of ERTMS will allow this work to be widely applicable in the railway industry.

This section starts with the introduction of the signaling context for our considerations of railway operations, and then it will present a brief overview of the existing regulatory framework towards putting forward the need of this work.

3.4.1 European Rail Traffic Management System

ERTMS is often cited as a major ‘European industrial project’ aiming to enhance cross-border interoperability by creating a single Europe-wide standard for railway signaling. It is composed of the European Train Control System (ETCS): a standard for train control and GSM-R: a GSM mobile communications standard for railway operations. Technically, it combines automatic train protection (ATP) and train control with the ability to enhance network capacity through more efficient traffic management.

From train driver’s perspective, modern signaling systems consist of an information system which relays relevant information necessary to ensure safe and timely operation. ERTMS is such a modern signaling system and a description of some of it’s components is given as follows. ETCS consists of elements of signaling, train protection system (ATP: Automatic Train Protection) and other core functions. ETCS requires standard hardware (on-board and track-side) and software components to function. To this end ETCS is divided up into mainly three different functional levels. The definition of the level depends on how the track and train are equipped. The driver machine interface and information displayed varies considerably between these levels. Thus, a brief description of these levels is given below, more details

can be found in [Schön et al., 2013].

3.4.1.1 ERTMS application levels

The ERTMS, depending on the level of implementation, is a partially ground (line-side equipments such as: Lineside Electronic Unit, Euroloop, Eurobalise, etc.) and partly on-board (Eurocab) signaling system. Technically speaking there are five different modes of ERTMS: Level 0, 1, 2, 3 and NTC/STM. Level 0, is used to handle cases where either the train or the track is not equipped with ERTMS/ETCS equipment, hence the default signalling system is to be used. NTC/STM is the case where train is under the supervision of a National Train Control (NTC) system which is interfaced with the use of STM (Specific Transmission Module) to the train's ETCS system. Levels 1, 2 and 3, are the core ERTMS application levels. These levels express the possible operating relationships between the signaling system and the train. These levels are of specific relevance to the present thesis hence, a brief explanation is given as follows:

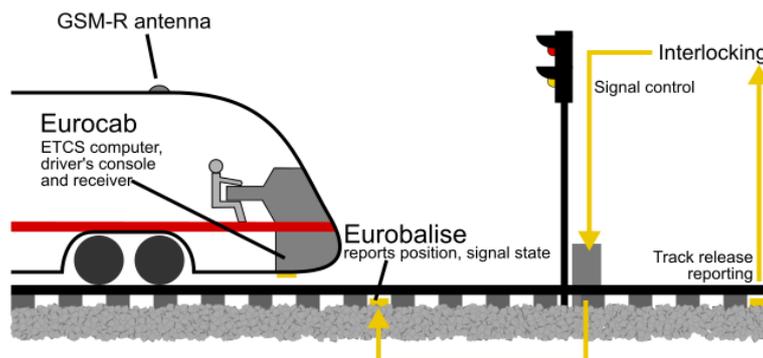


Figure 3.4 – Architecture of an ERTMS/ETCS level 1. Source: [Wikipedia, 2017]

In *ERTMS level 1*, a one-time transmission of information from track to train is done by the Eurobalise on the track. Track-side signals guide the driver on the route as shown in Figure 3.4.

In *ERTMS level 2* there is a bidirectional transmission of continuous information provided by GSM-R. The balises (also known as Eurobalise in the context of ERTMS) are used to enable the train to determine its location. Track circuits are used to detect the zone occupation of the train (Figure 3.5). All this information is relayed to a module called Radio Block Center (RBS), which then re-transmits via radio (GSM-R) signaling related information to the train. In this case track-side signals are no longer required because the relevant information is displayed directly in the cabin of the driver (Eurocab).

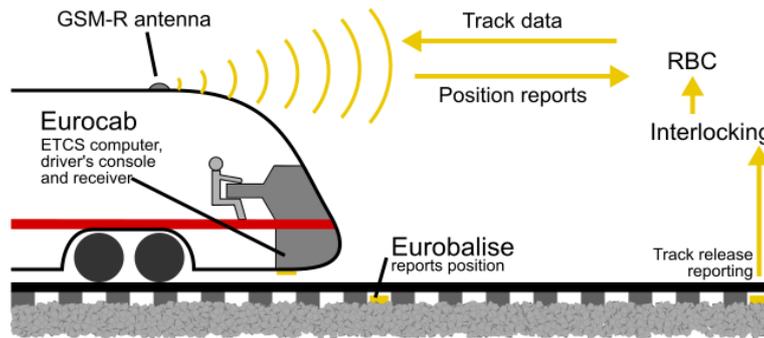


Figure 3.5 – Architecture of an ERTMS/ETCS level 2. Source: [Wikipedia, 2017]

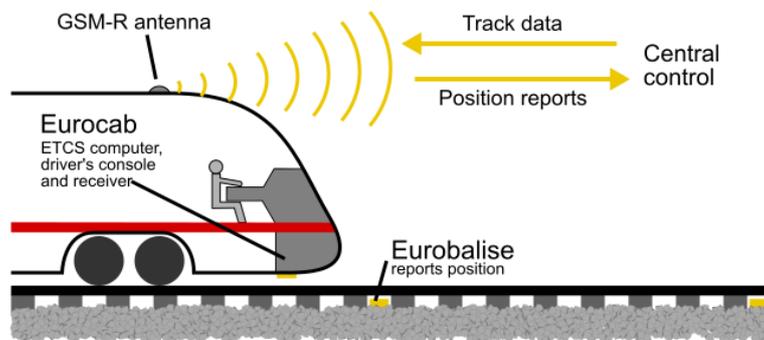


Figure 3.6 – Architecture of an ERTMS/ETCS level 3. Source: [Wikipedia, 2017]

ERTMS level 3, is based on the moving block concept similar to CBTC (Communication Based Train Control), the on-board system verify train's integrity based on its location, without the need of a detection systems (track circuits). In this case the balises are used by the on-board systems to update the location of the train by transmitting the information using GSM-R (Figure 3.6).

3.4.1.2 ERTMS/ETCS braking curves and train driver DMI

One of the principal tasks of a train driver is respecting the signaling. In a cab-driving context (that is the case for ERTMS Level 2 and 3) most signaling related information is delivered to the train driver in the cabin. This is done using the DMI - driver machine interface. On the back-end of this DMI, one of the core inputs to the ATP function are the so-called 'braking curves'. These curves form an important component of the safety relation between the train driver and the on-board systems, both are actively involved in ensuring a train's safe operation.

We cite [European Railway Agency, 2016], a document from EUAR, which

explains the concepts behind the ETCS braking curves. We are interested in the aspect of a human operator, hence we describe how these braking curves are created and subsequently analyze the *advising the driver* function of the braking curves. An excerpt from [European Railway Agency, 2016] is given as follows:

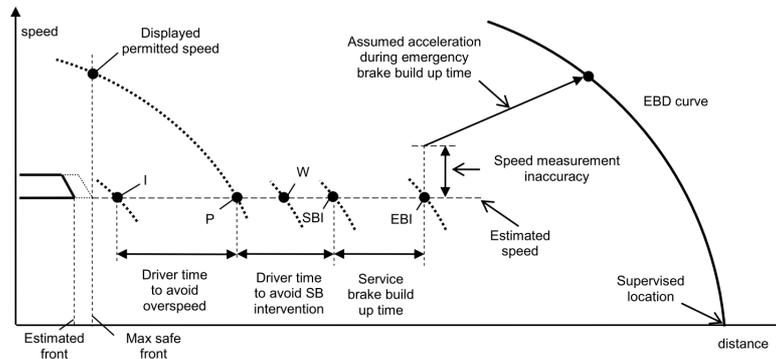


Figure 3.7 – Overview of the ETCS braking curve and its related supervision limits. Source: [European Railway Agency, 2016]

" In addition to the parachute functionality, ETCS provides the driver with advance information related to braking. Its purpose is to assist the driver and to allow him to drive comfortably, by maintaining the speed of the train within the appropriate limits. Therefore the ETCS on-board calculates in real time other supervision limits: Indication (I), Permitted speed (P), Warning (W) and Service Brake Intervention (SBI) (only if the ETCS on-board is designed to command itself the service brake). They consist of locations that, when crossed by the train, will trigger some information to be given to the driver through appropriate graphics, colors and sounds on the Driver Machine Interface. These locations are defined in order to:

- For the *I* [indicated] supervision limit: leave the driver enough time to act on the service brake so that the train does not overpass the Permitted speed, when this latter will start to decrease. Without the indication it would not be possible for the driver to perform a transition from ceiling speed supervision to the target speed supervision without over passing the Permitted speed *P*.
- For the *P* supervision limit: in case of overspeed, to leave the driver an additional time to act on the service brake so that the train will not overpass the point beyond which ETCS will trigger the command of the brakes.
- For the *W* supervision limit, to give an additional audible warning after the Permitted speed has been overpassed.
- For the *SBI* supervision limit, to take into account the service brake build up time so that the *EBI* supervision limit is not reached after the command by

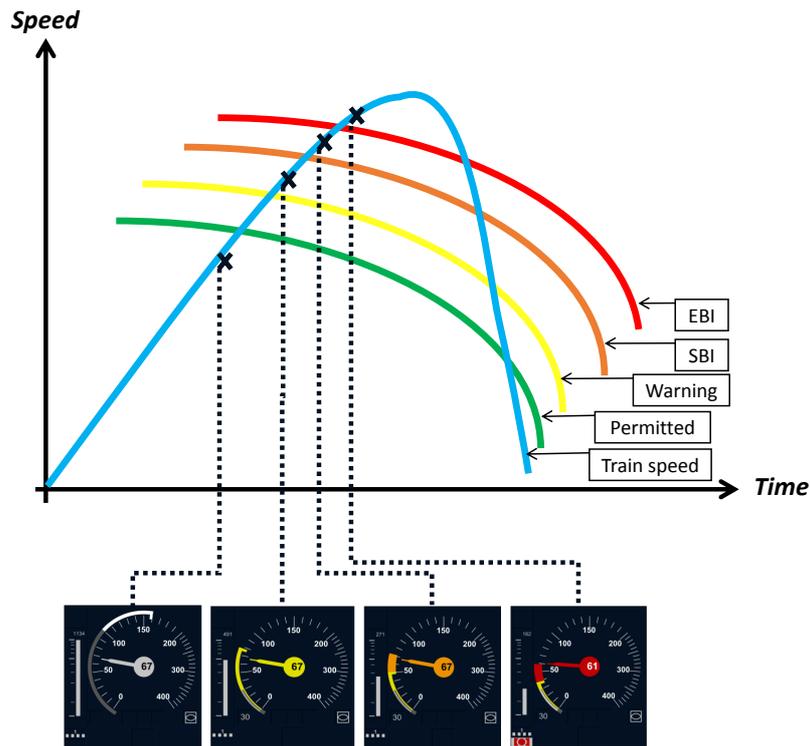


Figure 3.8 – ETCS DMI’s speed displays different representations based on the train speed vs. braking curves (colors represent the different speed curves also indicated by the labels on the right.)

ETCS of the full service brake effort. The SBI supervision limit is optional and can be implemented on-board the train in order to avoid too frequent emergency braking, which can be damaging for both the rolling stock and the track.

"

Essentially braking curves are generated by the ETCS on-board system to protect the train against unauthorized movements. It also informs the driver with appropriate assistance to allow him to drive comfortably. The different braking curves are indicated to the train driver using different colors and displays on the DMI, these different displays are given in [Figure 3.8](#). To note that there is also a sound alert associated with some speeds (warning, SBI speed and EBI speed).

Furthermore they also are aimed to be fully harmonized, that is various functional parameters and values are fixed. These fixed parameters are defined in the ERTMS SRS [[UNISIG, 2012](#)]. Some of these parameters define the relations between these curves in terms of time and speed values. We refer to the baseline 3 for the following discussion and subsequent usage [[UNISIG, 2012](#)]. The speed parameters and how different braking curves are generate are illustrated in [Figure 3.9](#).

The parameters this work is concerned with are the ones which define the

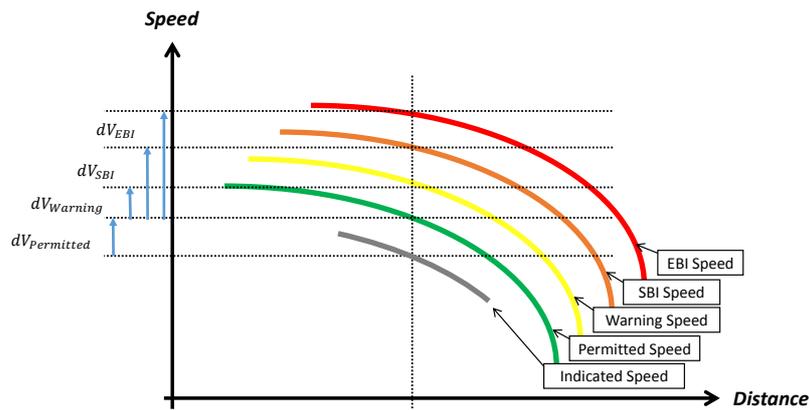


Figure 3.9 – Speed Curve parameters defined as ETCS fixed values Source: [UNISIG, 2012] (colors represent the different sped curves also indicated by the labels on the right.)

indications to the train driver. Their values and a brief explanation is given in [Table 3.1](#). These values (all the dV_i shown in figure 3.9) are defined as a pair of fixed values. For a given speed limit S , dV_{S_min} and dV_{S_max} are used by the on-board system based on the speed of the train.

Table 3.1 – Fixed values data and explanation for ETCS braking curves, also visible in [Figure 3.9](#)

Parameter re-fer Figure 3.9	Source text and explanation from [UNISIG, 2012]
$dV_{WARNING}$	Defined as Speed difference between permitted speed and Warning supervision limits. $dV_{WARNING_max} = 5km/h$ and $dV_{WARNING_min} = 4km/h$.
dV_{SBI}	For this a T_driver : “driver reaction time between Permitted speed supervision limit and SBI”, a fixed value is defined as 4 seconds, $dV_{SBI_max} = 10km/h$ and $dV_{SBI_min} = 4km/h$
dV_{EBI}	Speed difference between Permitted speed and EBI supervision limits, $dV_{EBI_max} = 15km/h$ and $dV_{EBI_min} = 7.5km/h$, we take the maximum value.
$dV_{PERMITTED}$	This is the difference between the indicated speed and the permitted speed. *Although this value is not explicitly defined, we assume it to be $5km/h$, since such a difference (a difference of 5) is seen for other values in this table. Note that this is an acceptable approximation, since passing indicated speed and driving at permitted speed is not a safety violation.

3.4.2 European regulations and human errors

The need to have a homogenized European railway network has opened a Pandora's Box of issues: technical agreements, signaling systems, operating rules, local non-signaling specific rules, etc., are some examples. On technical side of things, there are some railway specific standards published by CENELEC (French: Comité Européen de Normalisation Électrotechnique; English: European Committee for Electrotechnical Standardization) that are applicable to guide the development and certification of rail systems, some standards are listed below:

- *EN 50126* – The Specification and Demonstration of Reliability, Availability, Maintainability and Safety (RAMS)
- *EN 50128* – Communications, Signaling and Processing Systems – Software for railway Control and Protection systems
- *EN 50129*: Communication, signaling and processing systems - Safety related electronic systems for signaling.

They have among other objectives to guide the safety and reliability of railway systems. **The** standard for RAMS requirements of railways is EN50126 [CENELEC, 1999]. Among other guidelines, it gives an overview of what factors can influence the RAMS for railways as seen in [Figure 3.10](#). As an illustration we indicate the present work's context using a cloud shape to signify the fuzzy nature of a human in RAMS activities. In the current state of the norms, human factors are recognized to play a central role in RAMS assurances.

However, safety and reliability of a complex systems is a never ending endeavor, and secondly as visible in the interest behind systems like ERTMS, there is a need to facilitate the cross border operations in EU. Here, even though there are technical standards defined at the EU level, every country has independent safety authorities or regulators (EPSF for France, RSSB for UK) who have the final word on the safety considerations and risk analysis.

Some of the pressing issues, to be addressed are risk assessment methods and acceptance criteria. The European commission issued a mandate to European Union Agency for Railways (EUAR), for creating CSM Common Safety Methods for risk assessment [European Parliament, 2004]. It gives the following definitions:

- **“Common Safety Targets (CSTs) means the safety levels that must at least be reached by different parts of the rail system (such as the conventional rail system, the high speed rail system, long railway tunnels or lines solely used for freight transport) and by the system as a whole, *expressed in risk acceptance criteria.*”**

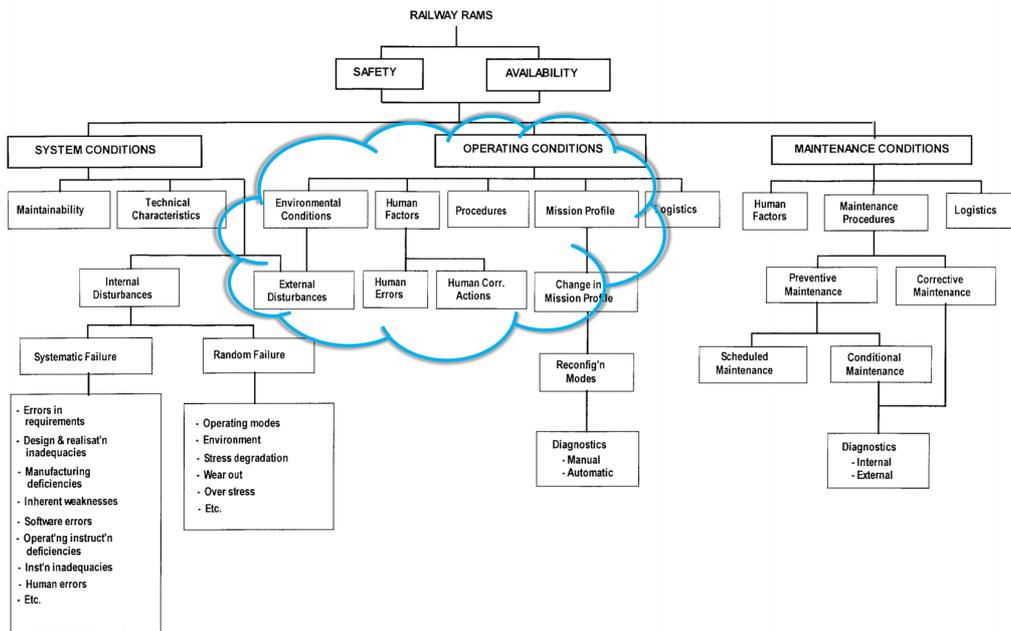


Figure 3.10 – Factors which can have an impact on RAMS of railways as per EN50126 [CENELEC, 1999]

- Where “*expressed in Risk Acceptance Criteria*” (RAC) is further defined as “*individual risks relating to passengers, staff including the staff of contractors, level crossing users and others, and, without prejudice to existing national and international liability rules, individual risks relating to unauthorized persons on railway premises, and societal risks*”.
- and finally the global context in which these targets and acceptance criteria are defined is the CSM. “*CSM are the means and methods to be developed to describe how safety levels and achievement of safety targets and compliance with other safety requirements are assessed.*”

CSM regulation was first proposed in 2009 and has since gone through multiple changes. EUAR has carried out multiple studies under the development of CSM [European Railway Agency, 2009] [European Commission, 2011], and for human factors risk in [Pickup et al., 2013] [Mowitz and Kecklund, 2013]. More specifically, a study [Det Norske Veritas, 2010] titled *Risk Acceptance Criteria for Technical Systems and Operational Procedures* aimed to identify the state of CSM and RAC in other domains (aviation, nuclear, maritime, etc.) in EU. They concluded that bow tie Quantitative Risk Assessment (QRA) is the preferred method to demonstrate compliance with risk acceptance criteria and “[in] these methods human reliability is usually included explicitly”.

In order to further develop the CSM and RAC other studies were carried out,

we discuss some of them mainly related to human factors. A survey [Kecklund et al., 2013] focused on the human factors, titled *Study on the Assessment and the Acceptance of Risks Related to Human Interactions within the European Railways* present the current state of how human factors related risks are addressed in EU. It concludes: “*even though the respondents participating in the survey performed some types of risk assessments as related to human interaction, they did not necessarily use any established human factors technique for this*”. They further conclude that the respondents “*work systematically even though they might not necessarily use specific, widely used techniques from the human factors or human reliability domain.*”, and further “*no special techniques for risk assessment or for the analysis of human interactions are used.*” Thus, even though the industrial actors address human factors in some way or the other risk assessment related to human factors, the current state of regulations has been accused of no explicit considerations of human factors, compared to the focus on the technical side of things [Di Grazia et al., 2014].

3.5 Conclusions

This chapter has introduced the context of this thesis – human errors and transportation system of system was also presented.

The particular problem of human error and some basic notions were also briefly discussed. There are various approaches to address human errors and the general notions of how to integrate HF in risk assessment approaches. The need to systematically include HF has been identified to increase cross-border EU-wide interoperability of railway traffic. Further, the complexity of a transportation system of system like railways is only increasing. Thus in the EUAR-led studies there is a frequent identification of human reliability analyzes.

Thus, the questions raised are what approach to follow to be able to (1) provide specific HF focused approach to account for their complex and systemic nature, (2) to be able to integrate HF aspects in risk analysis, and (3) to provide a methodological approach to address the human factors-related safety of transportation SoS?

To be able to respond to these questions, we will begin by asking more questions. What is the current state of such methods in literature, and other domains?, what are the desired characteristics of such methods? and what fits best for railways applications? These questions will be addressed in the next chapter.

A survey on human error quantification and its application to railway transportation

Contents

4.1 Introduction	51
4.2 Quantitative HRA – variables, data and frameworks	54
4.2.1 Identification of the variables of a quantitative model	56
4.2.1.1 Human failure event identification	57
4.2.1.2 Performance Shaping Factors	57
4.2.2 Error modeling and sources of data	62
4.2.2.1 Expert data	62
4.2.2.2 Empirical and experimental data	64
4.2.3 Quantification frameworks: a classification	65
4.2.3.1 Multiplier-like	66
4.2.3.2 Expert focused	70
4.2.3.3 Probabilistic graphical model-based	74
4.2.4 A summary and comparison	78
4.3 HRA in rail transportation	81
4.3.1 Variables of a quantitative framework for railway	82
4.3.1.1 Human failure event for rail operation	82
4.3.1.2 PSFs for railway and related works	85
4.3.2 Some quantitative considerations of human errors and frameworks in railway	91
4.4 Discussion towards a complete railway HRA methodology	94

4.5 Conclusions 96

4.1 Introduction

Human Reliability Analysis (HRA) finds its origins in the early 1970s; from where it started being included as an integral part of PRA (Probabilistic Risk Assessment) of nuclear power plants [US NRC Regulation, 1975]. Over the years, a large number of HRA models were proposed, developed and used for a quantitative and qualitative analysis of human actors involved in the operation of a the nuclear domain [Blackman and Gertman, 1994] [Spurgin, 2009]. First uses of HRA in the PRA of nuclear reactors were for classic control room and paper procedures. The contribution of human factors in system wide risk was initially underestimated to be around 15%, subsequent work placed in the range of 60-80%. First uses were done for classic control room and paper procedures. They adapted existing HRA models in order to analyze tasks and procedures with extensive use of data from simulator [Bot, 2010]. In general, the family of HRA methods and technique can be defined as follows: “. . . HRA is the use of systems engineering and behavioral science methods to evaluate the interaction between humans and the system, including the identification, qualitative analysis, and quantitative analysis of human actions, so that the impact of these actions on overall system reliability and their contribution to risk can be understood and managed.” [Chandler et al., 2006a]. After some early critiques [Dougherty, 1990] and other some changes were seen in the philosophy of newer HRA models. The integration of some such newer concepts were made and we had the so-called second generation of HRA models. The development of second generation tools began in the 1990s and is still on-going. However, a clear "ideal" classification is not evident; there are plenty of approaches each with its own merits. This makes it difficult to chose the aspects to be included or excluded. The [Boring, 2007] four Cs the 'classification factors' – *Cognition, Context, Commission and Chronology* – present a starting point, as to what a second generation model looks like. However, even with this classification, it is not possible to clearly determine the suit-ability or quality of neither a particular HRA method nor a generation thereof. To further complicate the issue it has been shown that several HRA methods may not give similar results and therefore cannot be compared [Reer, 2008]. Since, then the organizations have subsequently modified their models and methodologies to newer system designs and advances in the field of HRA.

For a cross-domain application a detailed critical analysis of the underlying notions are needed. Some of them are discussed below. Human reliability at first step confronts the problem of its own definition. First, it can be defined as technical reliability, i.e. the ability of a human component to realize its allocated functions successfully, in given operational condition and during an interval of time. A

measurement of this ability is usually the probability of success. However, this definition is not sufficient [Vanderhaegen, 2001], the human reliability is not static but evolves dynamically regarding learning effects and cooperative activities [Vanderhaegen, 2011], and its assessment is rather multi-criteria than mono-criterion. It usually relates to tasks to be achieved by human operators instead of functions and to the characteristics of these tasks and of the human resources [Vanderhaegen, 1999b].

The tasks characteristics concern the constraints of the task achievement such as: task is recoverable; the task is interruptible; task is monotonous; task is repetitive; task is simple or complex; the task allocation is preemptive, etc. The human characteristics are the human constraints for achieving tasks. There are constraints such as: humans are seen as a whole component or are composed by separate sub-components; humans are overloaded or under loaded; humans are hypo vigilant; humans are not experienced, etc. As a common starting point, human reliability and human error can be defined in terms of the causes of human behavioral dysfunction and/or their consequences for the system. Most HRA methods are thus, risk assessment-based and or cognitive model-based methods. They assess or analyze the risks of human or system dysfunction due to human actions [Vanderhaegen, 1999a]. Over the years various HRA models have been proposed which address different aspects of human machine interaction, represents human error mechanisms, associated quantitative and qualitative data, all towards the aim of making the system design or operation safe.

The probability of success of the control of dissonances such as contradictory knowledge, knowledge discovery or affordances has then to be taken in account [Vanderhaegen and Zieba, 2014]. The probability of success of the control of new situations or of unprecedented situations forces human operators to apply so-called trial-and-error based behaviors, and to discover new knowledge or to adapt their current knowledge [Ouedraogo et al., 2013] [Vanderhaegen and Caulier, 2011]. Human reliability assessment can have then several sources of explanation [Vanderhaegen, 2010]: the assessment made by the designers of a given human-machine system, the assessment made by an industry that will employ people in order to operate on this system, and the assessment made by the users of such a system. Sometimes these assessments differ. The feedback of experience is then required in order to integrate the natural learning effects from human operators into the design process and to take into account the behaviors applied for controlling well-known or unprecedented situations. Joint prospective, retrospective and on-line approaches are useful in order to guarantee the efficiency of the human reliability assessment. Evidential networks or Bayesian networks can

then be suitable tools to support such assessment [Aguirre et al., 2013b], [Sedki et al., 2013]

The need to address human errors has led to various reviews and guidelines from different domain perspectives: a study for the nuclear domain [Bowie et al., 2015] to support nuclear regulatory authorities; a HRA review study for space applications in [Chandler et al., 2006b] and a PRA guide which includes aspects of HRA [Stamatelatos et al., 2011]; a critical review for managers in high reliability organizations in [French et al., 2011]; an overview of HRA techniques for manufacturing operations in [Di Pasquale et al., 2013], health industry [Health and Safety Executive, 2009], [Lyons et al., 2004]. Finally, an in-depth study on human error in road transport in [Salmon et al., 2005]. The work in [Mosleh and Chang, 2004] although done from for the nuclear domain lists some desirable characteristics of HRA models, notably applicability to other domains, a procedure for quantitative results and the need for a model-based approach.

Furthermore, criteria for risks related to human errors, and their assessment needs further work as recognized in the latest amendment to CSM (Common Safety Methods) [European Railway Agency, 2015b]. Furthermore, the study [Kecklund et al., 2013] of rail entities Railway undertakings (RUs), Infrastructure Manager (IM) and National Safety Authority (NSA) from 10 European countries concluded that *"even though the respondents participating in the survey performed some types of risk assessments as related to human interaction, they did not necessarily use any established human factors technique for this"* and further *"most of the responding RUs and IMs do not use any specific human factors techniques."* They further remarked that there is a *"need to increase the knowledge on risk assessment of human interaction within the European railway system and to further increase the exchange of information on this topic within the European railway community."* The European Union Agency for Railways (EUAR) has carried out various human factor, and related risk analysis studies carried in the last few years [Pickup et al., 2013, Det Norske Veritas, 2010]. Also as elaborated in [Rail Safety & Standards Board, 2008] human factors considerations at various levels of railway operations are important.

Thus, for the rail domain the need for a dedicated method to assess human reliability was felt due to (i) an increasing involvement of humans in accidents while hardware reliability has steadily improved and (ii) the availability of very few methods to measure the risk of human towards the safe operation of the system. In addition, a regulatory need is also felt, as stated risk analysis related to human interactions and their evaluation, need to evolve and be recognized by regulatory and operational authorities.

Thus, towards a rail-HRA methodology, the rest of this chapter is structured as follows,

- The first part presents a state-of-art of quantitative HRA (identification of errors, assessment: qualitative, quantitative, etc.) to identify the recurring notions and good practices.
- It focuses on the quantitative HRA, by presenting a classification of mathematical frameworks. It also discusses some recent developments, such as: the use of probabilistic graphical models (Bayesian networks, etc.). It ends with a comparison of these frameworks and their respective methodologies.
- The second part focuses on the railway domain. A similar structure is followed to discuss the previous works for the railway domain.
- Some PSF and railway related works are discussed, and followed by a proposition of a PSF list for railway operations.
- Some related works and initial propositioning for a quantitative HRA for railway domain are discussed.
- Finally, we discuss the challenges that remain to be addressed, and some possible solutions.

4.2 Quantitative HRA – variables, data and frameworks

Most initial HRA methods provide a quantitative technique aimed at identifying the probability of occurrence of human error, known as Human Error Probability (HEP). Human reliability and human error can be defined in terms of the causes of human behavioral dysfunction and/or their consequences for the system. Most HRA methods are risk assessment-based or cognitive model-based methods. They assess or analyze the risks of human or system dysfunction due to human actions. Over the years various HRA models has been proposed which address different aspects of human – machine interaction, take in different human error philosophies, employ their respective methodologies, associate quantitative and qualitative data all towards the aim of making the system operation safe.

Given a context of operation (environment, objectives, etc.) HRA model provides a framework, to predict human performance towards system-level risk assessment. Human Failure Events (HFEs) and Performance Shaping Factors (PSFs) or some

variations thereof, are the basic units of an HRA analysis [Spurgin, 2009]. This section focuses on the quantification. It first describes the variables – the basic units on which quantification will be performed, followed by the source of data and finally the mathematical frameworks. Some well-known methods will be used to illustrate the underlying notions. It is concluded with a table of the HRA methods which employ similar frameworks. In general, when talking about *analysts*, this work refers to the users of an HRA methodology, or actors who perform the safety analysis (PRA, etc.). When referring to *experts* this work considers domain experts (e.g. expert in rail operations, or in human factors, etc.). An expert if involved in the analysis process can take the role of an analyst, offering their expertise to the application process.

Taking a PRA-prospective or similar final objective, allows clearly defining what a quantitative HRA model should focus on. The work in [Mosleh and Chang, 2004] proposes some desirable characteristics of what HRA methods *should enable* us to do:

1. identify human response (errors)
2. identify causes of errors to support development of preventive or mitigating measures
3. estimate response probabilities (error probabilities)

And more importantly it lists some guiding characters of such models:

1. include a systematic procedure for generating reproducible qualitative and quantitative results
2. have a causal model of human response with roots in cognitive and behavioral sciences
 - elements (e.g. PSFs) that are directly or indirectly observable
 - a structure that provides unambiguous and traceable links between its input and output
3. be detailed enough to support data collection, experimental validation, and various applications of PSA. Data and model are two tightly coupled entities.

A complete HRA method is often said to comprise of three elements: Identification, Modeling, Quantification. ATHEANA and THERP are often cited as being complete HRA methods [Barnes et al., 2000] [Bowie et al., 2015]. Thus, as seen in [Mosleh and Chang, 2004] and other discussions [Kyriakidis, 2013] a **complete HRA methodology** is comprised of the following entities:

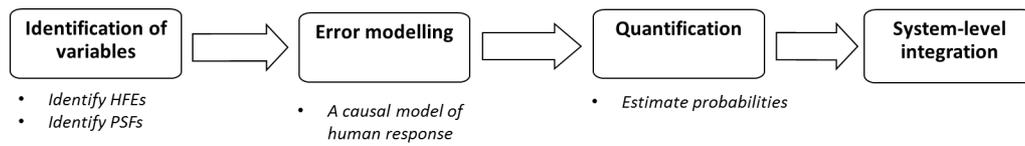


Figure 4.1 – Steps of a HRA quantitative framework

1. allows identification of: human response (errors) and the factors that affect the performance (PSF) (identification of variables: [Figure 4.1](#))
2. identify and model causes of errors: a causal model (error modeling [Figure 4.1](#))
3. allows performing the above analysis in both qualitative and quantitative manner
4. estimate probabilities: has a quantitative part (quantification [Figure 4.1](#))
5. should have a guide detailed enough to support data collection, experimental validation

The following discussion will focus along these lines. To guide the reader we propose to use the schematic in [Figure 4.1](#). A color code will be used to illustrate the differences between different methods. A box as seen in [Figure 4.1](#) will be colored green if the method under analysis includes guidelines on that step, and colored red if it does not. The central points of analysis are the objectives of a quantitative HRA. To note that the discussion does *not aim to present the usage* of such a framework. The main objective here is to illustrate the different characteristics of how each step is performed. However, these steps are often for similar objectives. Thus, the diagram [Figure 4.1](#) will be used to guide the reader on how different methods achieve the same objectives.

4.2.1 Identification of the variables of a quantitative model

HRA is often broadly classified into *generations*. HEART (Human error assessment and reduction technique), THERP (Technique for Human Error Rate Prediction) [[US NRC Regulation, 1975](#)] among other, are regarded as first generation. Models which include a particular focus on environment/context, some based on cognitive models like ATHEANA (A Technique for Human Event Analysis), CREAM (Cognitive Reliability and Error Analysis Method) are considered to be a second generation. Experts believe that human factors are not to be considered in isolation: environment, cognitive state and limited experimental data among other ambiguities are

needed to be addressed [Mosleh and Chang, 2004]. This for one, states the need to establish and focus on the variables of a quantitative model, which will form the basis of a quantitative framework's objectives.

Thus, the identification, often resulting from a qualitative analysis forms one of, if not the most crucial component of any HRA method. Evidently it is not limited to the HFEs (errors), but also PSFs.

4.2.1.1 Human failure event identification

Task analysis is normally a second step, that is where the system/human actors under analysis is already known. In general, when a system is more vulnerable to human error, a larger score and more comprehensive analysis is needed. This section omits the first step, since it is generally pre-HRA application. The definitions of some basic terms are presented before entering into details. **HFE - Human Failure Event** is a basic event, identified as part of a PRA (Probabilistic Risk Assessment) or a functional FTA (fault tree analysis) or task/procedure analysis. From a PRA-like analysis, an HFE is the failure of a function, system, or a component resulting from a human action/inaction. The definition and the context of an HFE are linked to the PRA analysis's point of view.

HFE, when identified can be a singular task or also involve multiple tasks. From an analysis point of view these tasks are sometimes grouped in terms of their generic characteristics. This generic grouping of variables is also found in some other models, for a task/HFE. In some models, nominal HEP defines a standard error rate for a certain task type. A PSF multiplier then increases or decreases that nominal error probability value. This nominal value depends on the type of task viz. task types in SPAR-H [Whaley et al., 2011] and Generic Task Types (GTTs) in HEART [Williams, 1985], etc. More details on these notions follow in the next sections.

4.2.1.2 Performance Shaping Factors

Performance Shaping Factors (PSFs) encompass influences that enhance or degrade human performance. PSFs are an integral part of the modeling and characterization of human reliability. These factors have been called by different names depending upon the method used: Performing Influence Factors (PIF), Influencing factors (IF), Performance Affecting Factors (PAF), Error Producing Conditions (EPC), Common Performance Conditions (CPC) etc. [Kim and Jung, 2003]. The definition to be considered remains the prerogative of the point of view taken by the model. However, most of them have similar connotations and are widely accepted by the

community.

Performance Shaping Factors lists – More specifically, PSFs allow the consideration of human’s own characteristics along with the context and environment which affect human performance in a negative or positive manner. For HRA objectives, in most cases focus is on negative influences. They allow an identification of contributors leading to a human error, providing a systematic basis for quantifying those contributors [Boring et al., 2007], [Boring and Blackman, 2007]. Thus, most HRA methods provide a list of factors that the analyst (user) can use to perform the analysis.

Some examples of PSF lists are given here to illustrate. The PSF list that is provided with the CREAM HRA model is given in [Figure 4.2](#). Such lists generally provides the PSFs, the states that a PSF can be assigned, and the effect it can have on human performance. The states are generally self explanatory in terms of their effect on human performance, e.g. an *inadequate* available time has a negative effect on human performance. Further, the eight PSFs used in SPAR-H [Gertman et al., 2005] are given in the [Figure 4.3](#). As it can be seen both models propose similar PSFs, and their associated states.

Performance Shaping Factors assessment – In order to define the degree of influences each PSF is accompanied by a certain number of *states or possible values*. In this section the discussion is limited to a qualitative assessment, the quantitative data associated with these states is discussed later, in the section ???. These values of PSFs are known as factor ratings [Podofillini and Dang, 2013], or rating scales [Gertman et al., 2005] or qualitative quality descriptors [Spurgin, 2009]. This work simply refers to them as states or levels. These states are discrete values a PSF can be in, and are important for HRA activities [Gertman et al., 2005]. Some use alternative phrasing based upon the need of the analysts or application (e.g. extra time for task load, etc.). Also the terms used for PSF levels can be modified depending on the focus. They can be factor specific (extra time for Time Load, high for training), or application specific [Rangra et al., 2015a]. For simplification we consider four levels (*good, nominal, poor* and *insufficient information*). They are defined (adapted from [Whaley et al., 2011]) as follows:

- *Good*: A PSF assigned this level is conducive to good performance, such that it reduces the opportunities for error, and thus, does not pose any safety issues.
- *Nominal*: It is assigned whenever a PSF is judged to support correct performance, but does not enhance performance (contrary to *good*) or make tasks

PSF	PSF State	Expected Effect on Performance Reliability
Adequacy of Organization	Very Efficient	Improved
	Efficient	Not significant
	Inefficient	Reduced
	Deficient	Reduced
Working Conditions	Advantageous	Improved
	Compatible	Not significant
	Incompatible	Reduced
Adequacy of HMI and operational support	Supportive	Improved
	Adequate	Not significant
	Tolerable	Not significant
	Inappropriate	Reduced
Availability of procedures/plans	Appropriate	Improved
	Acceptable	Not significant
	Inappropriate	Reduced
Number of simultaneous goals	Fewer than capacity	Not significant
	Matching current capacity	Not significant
	More than capacity	Reduced
Available time	Adequate	Improved
	Temperately inadequate	Not significant
	Continuously inadequate	Reduced
Time of day	Day-time	Not significant
	Night time	Reduced
Adequacy of training and preparation	Adequate, high experience	Improved
	Adequate, limited experience	Not significant
	Inadequate	Reduced
Crew collaboration quality	Very efficient	Improved
	Efficient	Not significant
	Inefficient	Not significant
	Deficient	Reduced

Figure 4.2 – Example of a PSF list: As proposed by the CREAM model [Hollnagel, 1998], and their states

PSFs	PSF Levels
Available Time	Inadequate time
	Time available is \approx the time required
	Nominal time
	Time available $\geq 5x$ the time required
	Time available is $\geq 50x$ the time required
	Insufficient Information
Stress/ Stressors	Extreme
	High
	Nominal
	Insufficient Information
Complexity	Highly complex
	Moderately complex
	Nominal
	Insufficient Information
Experience/ Training	Low
	Nominal
	High
	Insufficient Information
Procedures	Not available
	Incomplete
	Available, but poor
	Nominal
	Insufficient Information
Ergonomics/ HMI	Missing/Misleading
	Poor
	Nominal
	Good
	Insufficient Information
Fitness for Duty	Unfit
	Degraded Fitness
	Nominal
	Insufficient Information
Work Processes	Poor
	Nominal
	Good
	Insufficient Information

Figure 4.3 – Example of a PSF list: SPAR-H’s PSFs [Gertman et al., 2005] and the corresponding levels.

easier to carry out than typically expected.

- *Poor*: A poor level of a PSF is detrimental towards the accomplishment of an objective (leading to the occurrence of a human error).
- *Insufficient information*: If an expert/analyst does not possess sufficient knowledge to determine whether a PSF can affect a human's performance or if unable to choose among the other alternatives.

To note that, a state can be also PSF-specific, such as Inadequate Time for a PSF Time. Nevertheless, most of them aim to provide a similar basis as the definitions given above.

Reduced sets of PSFs are often used to represent situations of particular importance towards safety. Such considerations are more important for quantitative techniques, since they provide a reduced set of variables to work with (less variables, less data and more precise analysis). Characterization of such a situation indicates a safety critical context. Such a characterization of "critical context" is also employed in ATHEANA guidelines [Cooper et al., 1996]. They use the term error-forcing context (EFC). It is defined as "*particular combinations of performance shaping factors and plant conditions creating an environment in which unsafe actions are more likely to occur.*" EFC aims to present the experts or analysts potential interactions among the set of factors that are significantly different the usual influence of individual factors. Such collective nature of the PSFs (especially negative) needs to be considered rather than alone, at least when eliciting the experts [Forester et al., 2004]. These sets of PSFs reflect task and environmental characteristics, towards the sole consideration of human performance degradation. That is they are a collection of PSFs aimed at characterizing a context in which an error is more likely [Groth and Mosleh, 2012b]. Such an approach is also stated to be easier for usage. Further development leads to the concept of Error Contexts (ECs) from [Groth and Mosleh, 2012b]. Derived from empirical sources, ECs are defined as a construct which "describes certain combinations of PSFs that are more likely to produce human errors than the individual PIFs (similar to PSF) acting alone", also when compared to other combinations of PSFs. ECs are derived using data from multiple sources aiming to simplify the relations between variables of interest, in this case PSFs.

This dual nature of PSFs in characterizing accident contexts (frequency sourced reply to what went wrong), and accounting for human and contextual aspects (shaping human performance in general) make them an ideal candidate. Nevertheless, there is a strong case to be made for structured, hierarchical, well-defined and

exhaustive list of PSFs as proposed in [Groth and Mosleh, 2012a] for a detailed HRA analysis.

4.2.2 Error modeling and sources of data

Once all the variables are identified, the next step is to define the relations between the variables. At this stage we refer to the data which allows modeling the relationships between the PSFs and HFEs.

HRA domain suffers from a lack of data problem. Especially in the domain of risk and safety, no or very scarce frequency distribution or data samples are available. For an HRA context, even if they are, they are often linked to an application domain, with widely differing notions. This makes defining a frequentist probability for these cases inaccurate. Continuous probability distributions and similar representations are often used to make explicit representations and management of uncertainty [Aven and Zio, 2011]. Furthermore, integration of different incomplete sets of data, and from different sources therefore becomes necessary; concepts such as prior probabilities, data aggregation and beliefs can be employed.

4.2.2.1 Expert data

Expert-opinion elicitation is a formal process of obtaining information or answers to specific questions about certain issues that are needed to meet certain analytical objectives. Eliciting multiple experts and then combining or aggregating the data exist in many application domains and for multiple objectives [KIM and BISHU, 2006] [Knol et al., 2008]. Aggregation of data from multiple sources forms an essential part of a quantification objective, more so when the data is scarce, possesses uncertainties and is varied in terms of the nature of sources.

Two types of data aggregation can be identified [Budnitz et al., 1997] :

- Mathematical Schemes, in which expert inputs are combined using a mathematical formula. They include linear and logarithmic opinion pools, weights on the parameter values of underlying probability distributions, and Bayesian models.
- Behavioral Schemes, in which aggregation is accomplished through consensus or some type of qualitative argument. Most behavioral schemes are centered around some type of consensus process in which the group through either structured or unstructured interaction is given the task of reaching a consensus. They include Delphi methods and expert group interaction.

Often when using *mathematical aggregation* schemes dependence among the judgments of the experts is a critical concern. We cite [Hammitt and Zhang, 2013]: "*If multiple experts provide independent information, then an appropriate aggregate can be highly informative. Alternatively, if experts share much of the knowledge relevant to estimating a parameter value, the information contained in the union of their judgments may be little more than that contained in a single expert's judgment (in effect, each expert may report his idiosyncratic perception of a consensus).*"

Behavioral methods aims to generate consensus amongst experts by sharing of information and group discussions. Such methods are often resource intensive. On the other hand a wide variety of combination methods, algorithms exist, each of which allows managing different characteristics of the data, with their own hypothesis and mathematical formulations thereof. In [Ouchi, 2004], the authors discuss three expert data modeling approaches namely: Non-Bayesian Axiomatic Models (opinion pools, performance-based weight model), Bayesian Models and Paired Comparisons. They concluded that "*A general agreement appears to be that there is no single all-purpose aggregation method for expert opinion.*". Thus, the need of formal ways to combine expert judgment and empirical data remains an important issue [Mkrtchyan et al., 2016]. Since most HRA models are expert models, combination of knowledge taken from those experts is an aspect that needs attention and appropriate behavioral or mathematical methods should also be a part of a HRA process. In [Podofilini and Dang, 2013] an expert elicitation and combination process is given which aims to elicit probability distribution as estimates from experts. It is then used to build the quantitative model. These estimates are values of HEPs for specific PSF sets and levels, the so called conditional HEPs. They consider a cases of expert independence, which considers that (1) the experts are themselves independent and (2) each expert evaluates different HFES. Both of these conditions together satisfies the independence criteria. That is each expert is asked a different question. More cases of multiple experts and single question, are also considered as a case study.

As concluded in [Mkrtchyan et al., 2015] the need of formal ways to combine expert judgment and empirical data remains an important issue. Since most HRA models are expert models (as discussed later in this chapter) combination of knowledge taken from those experts is an aspect that needs attention and appropriate behavioral or mathematical methods should also be a part of a HRA process.

4.2.2.2 Empirical and experimental data

THERP handbook [Swain and Guttman, 1983] states: "*The necessity to rely so heavily on judgment (expert) is a regrettable state of affairs, but a start needs to be made, and this Handbook is a first step toward what is really needed . . .*". Thus the clear need to obtain empirical data was identified ever since the beginning.

There are various approaches, and with equally varying objectives to obtain human reliability data from simulators. Most of them aim to re-create real conditions and put human operators through multiple scenarios, observing the frequency of errors. Such an approach seems similar to the traditional approach for technical component testing. It is to determine a human error probability, in general given set of operational conditions. Since all possible scenarios and conditions cannot be simulated, therefore emergency (nuclear [Park and Jung, 2007]) or degraded conditions are preferred, i.e. the worst cases in terms of safety. This simulation data in raw form are used as databases which can be used to inform HRA activities [Park and Jung, 2007] and [James Chang et al., 2014].

Previous usage of simulators for human reliability activities have seen a variety of objectives for example, human reliability data collection, analysis of scenarios, validation of HRA models [Shirley et al., 2015]. There have been some general approaches which aim at gathering data for HRA purposes. In particular, a recent example of an extensive data collection activity [Lois et al., 2009]. This study aimed at collecting data from simulator runs - raw human performance data, towards evaluating the predictions of HRA methods.

In [Groth et al., 2014] the authors present a Bayesian methodology to update HEPs from existing methods using simulator data. For example, they have used data from the HERA database [Groth, 2009]. The HERA database, as detailed [Groth, 2009] "*...contains a detailed time line of sub-events, i.e., the successes and failures of hardware, human tasks and organizational elements.*". A sub-event is a single human task, equipment actuation or failure, or external state that occurs during an event. Further, for each sub-event there is an indication of the PSFs levels (a PSF was *adequate*, *less than adequate*, or if no details are available *nominal* or *indeterminate*). Such data gives at the very least a conditional data on $P(PSF/error)$. This data can be used to obtain probabilities values towards a conditional/subjective representation of the relation between HFEs and PSFs.

Among railway specific works [Qiu, 2014] present the usage of a rail traffic supervision simulator to obtain HEPs. The data obtained from the simulator are detection time, rate of correct detection, rate of false detection, and rate of non-detection. Their final objective was to obtain HEP by implementing a mix of

probabilistic graphical models and other HRA methods. In the context of ERTMS and train driving in particular, the work in [Rachedi et al., 2012] aims to evaluate a train operator's (driver) state. They characterize the state of a driver from data from experimentation on an ERTMS driving simulator, towards the objective of detecting driver's drowsiness or nervousness. These indicators are computed using non-intrusive data collection - notably from the speed curves. A more system-level human factors-oriented work is seen in [Belmonte et al., 2011], also on railway supervision application, they aim to evaluate the impact of controllers on the global safety of rail system. Their focus is seen on the casual chain of events, e.g. inadequate monitoring strategy leading to late detection leading to pressure on the diagnostic operation. The data measured/obtained are detection time, number of correct/incorrect detection, actions, etc. However, there is a feedback in the form of ergonomic enhancements, which follows more closely the notions of HF than HRA.

In [Musharraf et al., 2014] the authors present a methodology to collect human performance data for Bayesian network modeling applied to offshore oil rig operations. In their experimental set-up they considered three PSFs, and each PSFs had two possible states. A scenario was defined with each possible combination of the states of a PSF, i.e. 2^3 scenarios in total. For each of the scenario data was collected on some objective criteria (time-based, etc.) to build the conditional probability of the relations between the variables. Their use of scenarios with a certain states of PSFs, and a post-simulation evaluation of the objective criteria, gives data in a conditional form. This conditional data is adapted to be modeled using Bayesian models. Further, they gathered data to create models for a domain new to HRA (offshore oil rig); using simulators which are already used for training operators, such an approach minimizes cost and effort, and provides an empirical base for HRA modeling.

4.2.3 Quantification frameworks: a classification

The final variable of interest, for most HRA methods is a human error probability. In some cases this is driven by the global analysis (for example a PRA). The underlying mathematical framework can vary depending on the data, relations between variables and the usage of the model. This section propose to discuss three classes of quantitative HRA frameworks, mainly industrial-scale, and some newer research proposals. This discussion is limited to the more popular methods (many other have been proposed since late 1970s). For the newer proposals research work is referred due to the lack of industrially used methods. The main objective is to identify the nature of HRA quantitative modeling. This classification is a

hybrid approach to understand quantitative HRA modeling, it is not only aimed at understanding how a HEP is calculated, but what data is required and how the human reliability is modeled in the first place. To support this discussion, some considerations on the principal source and type of data needed or supplied with the model is also discussed. We focus on methods which employ PSF, this is the case with most well-known methodologies; further some specific techniques such as time-based methods are not considered. Most well known quantitative HRA models employ a variation of one of these three frameworks. Thus, the three classes proposed for discussion are:

1. Multiplier-based
2. Expert focused
3. Probabilistic Graphical Models

For each class a well-known method is chosen and explained to illustrate. The rest are not explored in detail but are listed at the end of section with some brief characteristics.

4.2.3.1 Multiplier-like

A multiplier-like framework relates the effects of a PSF on a nominal HEP ($NHEP$) value. Essentially it presents list(s) of data values, and to quantify HEP the analysts selects and multiplies the two selected values. It considers a nominal error probability of a human error, and then if a PSF is considered to be present in the scenario being analyzed, it's effect is considered as an increment in that nominal error probability. This is calculated using a multiplier value. Often both, the PSF multiplier value and $NHEP$ values are provided as a tabular form with a model's methodology. SPAR-H (Standardized Plant Analysis Risk-Human reliability analysis) [Gertman et al., 2005], [Whaley et al., 2011] is a well known quantitative HRA framework. The authors describe the source of these multiplier values in [Boring and Blackman, 2007]. SPAR-H is presented below to illustrate this first class of mathematical framework.

SPAR-H Step 1: Identifying and characterizing the HFE. For an HFE identified, this first step aims to characterize the HFE. This is done based on the type of the HFE. SPAR-H considers two types of HFE, either *diagnosis tasks* or *action tasks*. The underlying objective of such a classification is to assign nominal HEP ($NHEP$) values to an HFE. This $NHEP$ value is assigned based on the characteristic of the task type, for a HFE.

SPAR-H Step 2: Select and rate the PSFs from the given list of PSFs and select the multipliers This step is carried out with the analysts. An analyst analyzes the context and assigns ratings to the PSFs. These ratings are levels or states of a PSF. A pre-screening is required to understand if there is enough information to assign a value to a PSF. Secondly, if that PSF is present and influencing in that context. That is, first, if the analysts have enough data (about the HFE's scenario, procedure or plant conditions) *and*, second they can say that a given PSF will influence the operator's performance. If the response to both questions, in that case a PSF is considered in quantification - to be in either degrading (a poor), or enhancing state (good state). If none of these cases are identified, it is either considered to be having no effect (nominal state) or there is a lack of information (insufficient information state, essentially this PSF is not considered in quantification).

Once the level is assigned, SPAR-H provides multiplier values assigned for a state of a PSF, a detailed explanation and comparisons with other methods is given in [Gertman et al., 2005]. The table provided with SPAR-H's worksheets to the analysts is given in Figure 4.4. It gives the multiplier values for PSFs states for a task which can be characterized as an action task (a task type). We can see in Figure 4.4 that a state of a PSF *poor*, i.e. degrading human performance level increases the *NHEP* value multiplication with a value > 1 ; a *nominal* level has no effect ($NHEP \times 1$), and a *good* level reduces the *NHEP* multiplication with a value < 1 .

SPAR-H Step 3: Calculate PSF modified HEP. Once both the variables, the HFE (as *NHEP* value), and PSF (as the state it is in) are identified, the quantification is performed using the following equation:

$$HEP = \frac{NHEP \times PSF_{composite}}{NHEP \times (PSF_{composite} - 1) + 1} \quad (4.1)$$

where: *HEP* is the final human error probability; *NHEP* is the nominal HEP value; $PSF_{composite}$ is the combined multiplier effect of all the PSFs identified in the previous step.

SPAR-H Step 4 and Step 5: Dependency and cutoff value. It allows minor modifications in the HEP calculation based on the fact if the HFEs are dependent or not. In SPAR-H, HFEs are defined in the way that they are independent of one another. And finally, SPAR-H incites the analyst to ask the question "how small can an HEP become before it becomes unrealistic and unbelievable?". The value suggested by SPAR-H is 10^{-5} . These two steps are calibration steps for

PSFs	PSF Levels	Multiplier for Action	
Available Time	Inadequate time	P(failure) = 1.0	<input type="checkbox"/>
	Time available is \approx the time required	10	<input type="checkbox"/>
	Nominal time	1	<input type="checkbox"/>
	Time available $\geq 5x$ the time required	0.1	<input type="checkbox"/>
	Time available is $\geq 50x$ the time required	0.01	<input type="checkbox"/>
Stress/ Stressors	Insufficient Information	1	<input type="checkbox"/>
	Extreme	5	<input type="checkbox"/>
	High	2	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
Complexity	Insufficient Information	1	<input type="checkbox"/>
	Highly complex	5	<input type="checkbox"/>
	Moderately complex	2	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
Experience/ Training	Insufficient Information	1	<input type="checkbox"/>
	Low	3	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
	High	0.5	<input type="checkbox"/>
Procedures	Insufficient Information	1	<input type="checkbox"/>
	Not available	50	<input type="checkbox"/>
	Incomplete	20	<input type="checkbox"/>
	Available, but poor	5	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
Ergonomics/ HMI	Insufficient Information	1	<input type="checkbox"/>
	Missing/Misleading	50	<input type="checkbox"/>
	Poor	10	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
	Good	0.5	<input type="checkbox"/>
Fitness for Duty	Insufficient Information	1	<input type="checkbox"/>
	Unfit	P(failure) = 1.0	<input type="checkbox"/>
	Degraded Fitness	5	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
Work Processes	Insufficient Information	1	<input type="checkbox"/>
	Poor	5	<input type="checkbox"/>
	Nominal	1	<input type="checkbox"/>
	Good	0.5	<input type="checkbox"/>

Figure 4.4 – SPAR-H PSF states and respective multiplier values assigned for a state of a PSF for an action task[Gertman et al., 2005]



Figure 4.5 – SPAR-H quantification steps (boxes in green show the steps performed in SPAR-H)

the framework.

The application of SPAR-H is performed as seen in [Figure 4.5](#). In this figure the boxes in green signify that SPAR-H only deals with error modeling and quantification.

This class of methods is characterized by predefined lists of PSFs, multiplier values for the PSFs aiming for capturing the effects of PSFs. An HFE is generalized to assign predefined numerical values (*NHEP*). Some methods which employ a multiplier-like frameworks:

- THERP [[Swain and Guttman, 1983](#)] [[Boring, 2012](#)]
 - HEP calculation: Analyst selects the value: a *NHEP* (the nominal HEP value), from tables provided, and modifies it based on the task dependency.
 - Data source: table of *NHEP* values and other (PSFs table, etc). SPAR-H takes some of these values from THERP.
- HEART [[Williams, 1986](#)] [[Gibson et al., 2013](#)]
 - HEP calculation:

$$FinalHEP = BasicHEP \times \prod_i [(Effect_{EPC_i} - 1) \times State_{EFC_i} + 1]$$
 where: basic HEP, is associated with a GTT (generic task type) is a central value of HEP (see an excerpt from HEART [Figure 4.6](#)); $Effect_{EPC_i}$ or APOA is the proportion of effect of an error production condition EPC_i ; with its max effect given by $WF_{EPC(i)}$, the multiplier values for a EPC (a PSF)
 - Data source: table of *NHEP* values for GTT, and EPC multipliers as shown in [Figure 4.6](#).

Some other methodologies are, a railway specific method: Railway Action Reliability Assessment (RARA) [[Gibson et al., 2013](#)], based on HEART, discussed later. They have been readily adapted to various domains (aviation – CARA [[Kirwan and Gibson, 2007](#)], Petroleum [[Bye et al., 2016](#)]. NARA proposed in [[Kirwan et al., 2004](#)] as a refinement of HEART. Such methods are easy to use due to the limited human performance choices, the presence of guidelines and good documentation.

Generic task	Proposed nominal human unreliability (5th–95th percentile boundaries)	Error-producing condition	Maximum predicted nominal amount by which unreliability might change going from 'good' conditions to 'bad'
A	Totally unfamiliar, performed at speed with no real idea of likely consequences 0.55 (0.35–0.97)	1. Unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel	< 17
B	Shift or restore system to a new or original state on a single attempt without supervision or procedures 0.26 (0.14–0.42)	2. A shortage of time available for error detection and correction	× 11
C	Complex task requiring high level of comprehension and skill 0.16 (0.12–0.28)	3. A low signal-to-noise ratio	× 10
D	Fairly simple task performed rapidly or given scant attention 0.09 (0.06–0.13)	4. A means of suppressing or overriding information or features which is too easily accessible	× 9
E	Routine, highly practised, rapid task involving relatively low level of skill 0.02 (0.007–0.045)	5. No means of conveying spatial and functional information to operators in a form which they can readily assimilate	× 8
F	Restore or shift a system to original or new state following procedures, with some checking 0.003 (0.0008–0.007)	6. A mismatch between an operator's model of the world and that imagined by the designer	× 8
G	Completely familiar, well-designed, highly practised, routine task occurring several times per hour, performed to highest possible standards by highly motivated, highly trained and experienced persons, totally aware of implications of failure, with time to correct potential error, but without the benefit of significant job aids 0.0004 (0.00008–0.009)	7. No obvious means of reversing an unintended action	× 8
H	Respond correctly to system command even when there is an augmented or automated supervisory system providing accurate interpretation of system stage 0.00002 (0.000006–0.00009)	8. A channel capacity overload, particularly one caused by simultaneous presentation of non-redundant information	× 6
M	Miscellaneous task for which no description can be found. (Nominal 5th to 95th percentile data spreads were chosen on the basis of experience suggesting log-normality) 0.03 (0.008–0.11)	9. A need to unlearn a technique and apply one which requires the application of an opposing philosophy	× 6
		10. The need to transfer specific knowledge from task to task without loss	× 5.5
		11. Ambiguity in the required performance standards	× 5
		12. A mismatch between perceived and real risk	× 4

Figure 4.6 – An excerpt from the tabular predefined values given with the HEART methodology [Williams, 1986]: GTT’s NHEP values (left), and EPC multipliers (right).

Their off-the-shelf application is limited due to the limited choice of variables, and need work to modify and validate for a domain. The set of variables (predefined PSF list) and quantitative values (predefined tabular values) are often application focused. Since SPAR-H inspires from nuclear-domain specific data and previous models, questions can be raised on the validity of the pre-defined data for other domain of applications. At the very least, a SPAR-H like quantitative model should be backed by domain-specific data.

4.2.3.2 Expert focused

Such classes of frameworks are centered on constructing a quantitative model from an expert elicitation process.

To illustrate such models, we present, ATHEANA (A Technique for Human Event Analysis) [Cooper et al., 1996]. ATHEANA is a complete HRA methodology. It uses data from expert elicitation [Forester et al., 2004] for its quantitative part. Apart from the notion of HFE and PSFs, it employs EFC’s, as discussed in section 4.2.1.2. It is defined as a combination of PSFs, where human errors (unsafe actions) have a higher probability. These Unsafe Actions (UAs) are decomposition of an HFE in terms of the different ways an HFE can occur.

Unlike SPAR-H, quantification in ATHEANA’s quantative framework is one step (#8) out of total 9 steps of its complete methodology [Cooper et al., 1996]. Identification of the variables, and quantitative relevant part begins from Step 4, for ease of understanding this step is numbered as 1 in the following discussion. Original numbering of steps from [J. Forester et al., 2007] is given in parentheses (*). ATHEANA’s quantification related steps are briefly discussed below:

ATHEANA Step 1 (4*). Define the Corresponding HFE. Having identified the context of analysis, the purpose of this step is to identify the HFEs that need to be

analyzed. In addition, associated to this HFE they ask to identify an UA (unsafe action). A UA is defined as a decomposition of an HFE in terms of the different ways an HFE can occur

ATHEANA Step 2 (5*). Identify and characterize PSFs. This step is where the analysts identify the PSFs. Similar to SPAR-H, a focus is found on the PSFs which might contribute to performance. ATHEANA provides a supporting list of factors: training/experience, procedure, HMI quality (Availability and clarity of instrumentation, ergonomic quality), time requirements, Workload, Time Pressure, and Stress, etc. In total this list contains 16 PSFs. This step is followed by determining the positive and negative influences (i.e. states of these PSFs).

ATHEANA Step 3 and 4 (6 and 7*). Identify scenarios and Potential for Recovery. The step involves identifying different possible scenarios - "deviation scenarios" for which the quantification of HFE's will be carried out. A focus on the PSFs most negatively impacting in the identified scenarios is also done in this step. A screening of the scenarios is done next. This screening aims to identify if recovery is possible (with a high likelihood) for the HFEs for a scenario. If yes, the concerning scenario is not analyzed further.

ATHEANA Step 5 (8*): Quantification of HEP. This step is where the quantification of an HEP for a HFE (or an UA) is carried out. The following equation is used:

$$P(HFE|S) = \sum_i P(EFC_i|S) \times P(UA|EFC_i, S) \quad (4.2)$$

Where $P(HFE|S)$ – Probability of an HFE given a scenario; $P(EFC_i|S)$ – Probability of i th EFC given a scenario and $P(UA|EFC_i, S)$ – Probability of an UA given the EFC and the scenario; S represents the collection or series of events under analysis.

According to ATHEANA's expert elicitation process [Forester et al., 2004] states that this equation is "*not a mechanistic calculation.*" That is to say it provides concise representation of the data needed from the experts, "*it alerts us the need to examine a wide range of EFCs, given a particular UA associated with an accident scenario (i.e. S).*". It is just there to guide the experts in the elicitation process. A group-based elicitation process is proposed. After multiple sub-steps of explanation and discussion, the experts are asked for a probability distribution of the *HEP*. This is a two-part process. The first part asks experts to give reasons as to why an HFE might occur, a justification such as: "*The action will be (easy, hard, extremely difficult, etc.) for the operator if that because. . .*". Second part is giving experts a set of calibration

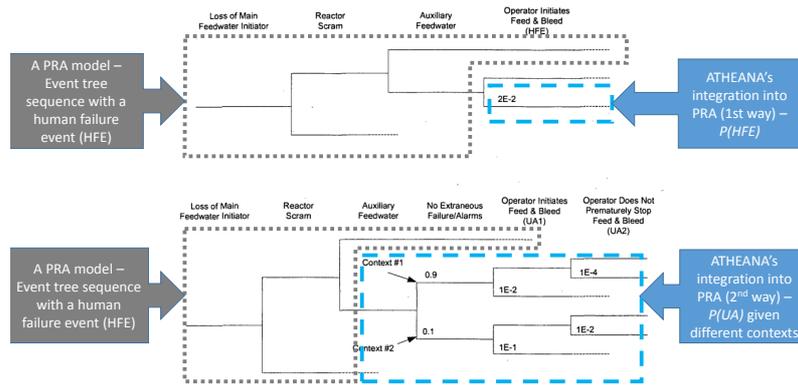


Figure 4.7 – Integration of ATHEANA’s results in a PRA model using two possible ways, adapted from [J. Forester et al., 2007]

points, also used in some other models as qualitative descriptors (the operator is "Likely" to fail = 0.5, "Infrequently" = 0.1, and so on). Hence, essentially the HEP is directly obtained from the experts, although through a significantly long and systematic process.

ATHEANA Step 6 (9*): Integration into PRA ATHEANA’s last step provides some guidelines on how to integrate it’s results into a standard PRA process for system-level results. They list two ways to integrate ATHEANA’s results in the PRA model. **Figure 4.7** shows the two possible ways. The guide [J. Forester et al., 2007] gives an example event tree that is a part of PRA model of a nuclear power plant. In this model where an *HFE* is identified, and ATHEANA is used to analyze that *HFE*. Note that the PRA model can be a event tree or fault tree etc. The two ways are explained as follows:

1. The first is to maintain the original PRA modeling and HFE definition (the top part of **Figure 4.7**). The HFE is treated as a success or failure event, and it’s HEP is calculated from **Equation 4.2** and added to the event tree.
2. The second way is to “expand” the original PRA model. For example, here for an event tree more top events are added (for a fault tree more basic events can be added). In **Figure 4.7** the different EFCs and the HFE is broken down into UA’s are all integrated in the modified PRA model.

However, they do not explicitly state how to identity the impact of the HFEs/UAs on system-level risk in the ATHEANA methodology and the PRA process manages that, hence **Figure 4.8** shows the system-level integration colored green-red.

These steps of quantification as compared with the steps presented previously. As can be seen in **Figure 4.8** ATHEANA provides guidelines for all of the steps of human



Figure 4.8 – ATHEANA quantification steps (boxes in green show the steps performed in the ATHEANA methodology; half green-red shows that it is not performed in ATHEANA, but guidelines are provided)

error quantification (almost all the boxes in [Figure 4.8](#) are colored green). However, the *NHEP* or similar values are not provided in the guidelines, the numerical values are obtained from the experts. Thus, it specifies how to quantify but no (or only indicative) predefined values are provided. Hence, the Error modeling step in [Figure 4.8](#) is colored half red and half green. Most such methods rely on expert data to quantify, and the mathematical equations are given to guide the experts, not to be used directly by the analysts. Some other similar methods are given below:

- SLIM [[Embrey et al., 1984](#)]
 - Mathematical framework (to combine expert data only) $\log(HEP_j) = aSLI_j + b$; where $SLI_j = \sum NormalizedWeight(PSF_i) \times StatePSF_i$ where, SLI_j is the combined weighted obtained from the experts, a and b are empirically derived constants from success probability of two related tasks.
 - Data source: Expert judgment and combination
- MERMOS [[Bieder et al., 1998](#)] [[Meyer et al., 2007](#)]
 - Mathematical framework on scenario, expert data, and conditional probabilities. $P(\text{failure of the HF mission}) = \sum P(\text{failure scenario}_i) + P_{residual}$. i.e. as stated in [[Meyer et al., 2007](#)] "*The total probability of failure of the HF mission is defined as the sum of all probabilities of occurrence of all failure scenarios identified, plus the residual probability, representing possible unforeseen scenarios*".
 - Data source: Expert judgment (based on tests on simulators)

Some methodologies with similar quantitative frameworks are - Human Performance Railway Operational Index (HuPeROI) [[Kyriakidis et al., 2012](#)] (based on SLIM, discussed later). Some characteristics of such models are: An extensive focus on qualitative (objectives of the analysis, scenarios, EFCs-PSFs, HFES-UAs); extensive discussion and documentation to ensure confidence and repeatability, few or only indicative predefined numbers. Some advantages of such an approach are that it works with lack of empirical data (hence the expert elicitation); most

of them are complete methodologies and are qualitative heavy. Furthermore, ATHEANA offers an adaptable approach e.g. learning from retrospective analysis for prospective analysis.

ATHEANA is said to be a second generation HRA model [Boring, 2007]. However, the increased appreciation of the second-generation models comes at a price, i.e. firstly, increased resources needed to develop the models, gather the supporting baseline empirical data and extensive validation studies, [Lois et al., 2009] etc, and secondly, the effort needed in their application. The ATHEANA complete methodology (only the quantitative part is discussed here) is an example of a model "too complex and difficult to be applied" [Kyriakidis, 2013]. Further, expert-based methods are in-general resources intensive when applying. ATHEANA implementation guidelines state "*analysis requires a broad range of multidisciplinary knowledge: behavioral and cognitive science, the plant-specific design and PRA. Understanding of plant behavior (including thermal-hydraulic performance), understanding of the plant's operational practices (including procedures, training, and administrative practices), and generic and plant-specific operating history (including incident history, backlog of corrective maintenance work orders, and current workarounds).*" [Cooper et al., 1996]. Thus, this process can be time consuming and training is required to be able to apply the methodology. Further, they (also other expert models in general) are critiqued to have questionable accuracy and repeatability [Barry, 1997]; they are relatively complex and resource intensive (training and time are needed).

4.2.3.3 Probabilistic graphical model-based

As discussed in [section 2.1.1.4](#), probabilistic graphical models have seen an interest for quantitative HRA application [Mkrtchyan et al., 2015]. In HRA usage most such frameworks are at the level of quantification for specific applications. However, they can be integrated on top of existing models as a quantitative module. Integration of fault tree and Bayesian networks can be performed as presented in [Martins and Maturana, 2013], this allows an adequate representation of the human component and precise quantification of human reliability.

The use of such frameworks also aids the notion of an integrated risk analysis where different domains technical, human and organizational aspects can be analyzed at the same level. In [Duval et al., 2012] the authors present a BBN-based framework to rank the risks related to these factors. The flexibility of these frameworks were further explored in HRA-related modeling: by integrating a PSF-based approach and the assessment of barriers [Galizia et al., 2015]. The approach in [De Galizia et al., 2016a] to integrate non-deterministic mechanisms, as the



H

Figure 4.9 – A BBN model structure proposed in [Groth and Swiler, 2013] as the interpretation of SPAR-H’s quantitative model

authors describe “mechanisms that do not necessarily increase the probability of producing an undesired effect.”, what can be equated to the good effect of a PSF, is a testament to the usefulness of such frameworks.

The following steps discuss a simpler approach of how a probabilistic graphical model can be used, or has been proposed as a quantitative HRA framework.

Step 1. Identification of HFEs and PSFs. The HFEs and PSFs need to be defined and identified, for the given context of application. Such a framework can use this step from other methodologies. As discussed previously, SPAR-H’s Step 1 or ATHEANA’s Step 4 either of these step can be used to identify the variables of a model.

Step 2. Structure of a BBN model. BBNs for HRA applications follow a structure where the output node is generally the failure event (HFE) [Mkrtchyan et al., 2016]. This HFE is binary in nature, representing a failure and a success state. Subsequently, the value of interest for this HFE is, is the probability of the node being in failure state – an HEP. The intermediate nodes are generally PSFs for a flat (non-causal relationship) model. For some, [Groth and Swiler, 2013] this probability is interpreted as conditional probability of an HFE (or system level-inference such as an accident) on the PSF(s). That is similar to what is seen for ATHEANA in Equation 4.2. However here it is a mathematical formulation, where conditional data is needed. Furthermore, in the scope of HRA activities, PSFs are limited only to direct influences on the quantification; this makes the model flat rather than hierarchical. This approach although not exhaustive and rich, is common with many other BBN based models ([Mkrtchyan et al., 2016]).

In [Groth and Swiler, 2013] a BBN version of SPAR-H is proposed, as shown in Figure 4.9. The authors aim to present the expressiveness of the BBN framework, and this model goes to show that existing models (like SPAR-H in this case) can be modeled using a BBN. The reasoning behind why such a particular structure

is obtained is detailed in [Groth and Swiler, 2013]. Apart from a basic structure like this, proposals of models also are seen representing the following structure [Mkrтчyan et al., 2015]:

- PSF multi-level relations - representing relations between high level PSFs (MOF - management and organizational factors) to low level PSFs.
- PSF multi-level relations - multiple PSFs combining to form an error context (Figure 4.2.1.2).
- HFE dependence modeling relation between multiple HFEs (HFE dependence as used in THERP).

Thus, different structures are seen as representing relations between different variables. To note that very few of these have seen industrial strength applications (vs. SPAR-H or SLIM for example). Hence, more focus is found on the using the full expressiveness of the framework. However, at the very least, replication of the structure of an already existing framework is possible as seen in Figure 4.9, detailed discussion can be found in [Groth and Swiler, 2013].

Step 2. Building the relation between variables. Once the variables and arcs signifying the existence of a relation are identified, the actual relations need to be formally defined. In simpler modes such as Figure 4.9 these relations define the conditional probability (a PSFs influence) on an HFE's occurrence. Conditional probabilities are thus used to define PSF-HFE relations. These conditional probabilities are defined as CPTs (conditional probability tables). The CPTs express the probability of each node, given the states of its parent nodes (i.e. for each arc in Figure 4.9). In [Groth and Swiler, 2013], some manual assignments are also made. For example if *Available Time* (a PSF) is *inadequate* (a poor/degrading state of a that PSF) and *Fitness for duty* is *Unfit*, the final HEP is assigned the value of 1.0 regardless of the state of the other PSFs. They further report that SPAR-H's quantitative part Equation 4.1 can be used to generate the CPTs, automatically using the software *Hugin Expert*). A general simple example, of what such a relation might look like is given in Table 4.1.

Irrespective of the source they require a large amount of data to build the models, and is often the roadblock for most application domains with a lack of data. However, their usage offers various advantages, thus various approaches are proposed in the literature to go around the problem of requiring extensive data to build such models [De Galizia et al., 2016b]. They can be made to use data from varying sources and quantity to build model. Data can be sourced from experts

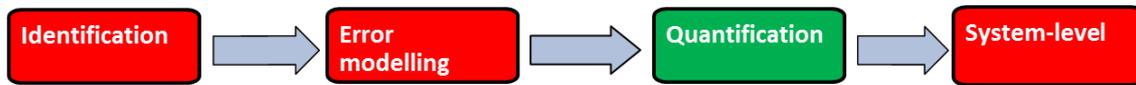
Table 4.1 – An example of definition of a relational data between the states of PSF and state of an HFE

PSF1	PSF2	PSF3	PSFn	HFE1	Belief / probability
poor	poor	poor	...	true	0.01
poor	poor	nominal	...	true	0.001
...
Insufficient info.	Insufficient info.	Insufficient info.	...	true, false	0

(questionnaires, interviews, etc. [Podofilini and Dang, 2013],), from empirical sources [Groth and Mosleh, 2012a] and other [Mkrtchyan et al., 2016] [Mkrtchyan et al., 2015] (qualitative data, etc.). It also offers a rich representation of uncertainty in modelling data, for example, expert confidence or multiple expert's agreement on the relation between two variables can also be integrated [Mkrtchyan et al., 2016]. Explicit representation and management of uncertainty, causal relations between variables (mainly PSFs and HFEs), make such frameworks a flexible and appropriate choice for quantitative HRA modeling. Furthermore, the graphical representation, combination of data, the ability to refine existing models also increases the usability of such models.

As can be seen in Figure 4.10, such an approach, from HRA standpoint does need integration or adaptation of a pre-existing model since it lacks, the identification of the variables and final integration of the quantification results at system-level. This is expected and as discussed previously, HRA methodologies focus on the human aspect and system-level analysis is done in a PRA. However, some methods do provide guidelines as to how to integrate or in some cases modify the PRA model to integrate a HRA's results. Further, in [Martins and Maturana, 2013] the parallels between the steps of THERP and a BNN-based methodology are presented, it can be useful to understand at what level and how such graphical models can be integrated in a quantitative HRA. The usage of probabilistic graphic models in HRA application is relatively recent, hence, at least from industrial application standpoint, there is a lack of complete HRA methodologies employing such frameworks, nevertheless some recent works present some concrete ways forward [De Galizia et al., 2016b] [Mkrtchyan et al., 2015].

Figure 4.10 – Quantification steps of a probabilistic graphic models HRA quantification framework



4.2.4 A summary and comparison

This section presents a summary of the pros and cons of the quantitative HRA classification previously. And summarize the methodologies they are part of, as previously discussed.

The first summary, in [Table 4.2](#) presents the pros-cons list of the three classes of quantitative HRA frameworks as previously discussed.

The second comparison in [Table 4.3](#) compares the internal components of the methods (steps, variables, data, etc.). This aims to summarize the HRA methodologies and complete the previous discussion.

The list below gives a resume of the variables that are identified in the various steps of different frameworks; the actors that are involved and the source(s) of data that are used. Thus, concluding the similarities, and some characteristics of the quantitative HRA irrespective of the framework selected.

Step 1: Identification of safety critical activities (HFEs) and contexts (PSFs)

- HFEs, UAs: (pre-HRA mostly) task, procedure or scenario analysis
- PSFs,EFC: Characterization/analysis of a context/situation
- Actor: analyst

Step 2: Error modeling

- HFE-PSF relation: Pre-existing data - empirical (accident analysis/simulator/other HRA) or expert elicitation (aggregation or consensus)
- Actor: expert(s)/analyst

Step 3: Quantification

- HFE/HEP quantification: using error model, other HRA models or expert data
- Actor: expert(s)/analyst

Step 4: System-level integration

Table 4.2 – Some advantages and disadvantages of the previously discussed three classes of quantitative HRA frameworks

Advantages	Disadvantages
Multiplier-like	
<ul style="list-style-type: none"> • They are easy to use. • Most such methods have been adapted to other domains. • They are used in most well-established and industrially used HRA methodologies. • A large number of adaptations with similar characteristics (PSF lists, <i>NHEP</i>, multiplier values, etc.) • If quantitative data is available, the model has been verified, and ease of usage is desired – this framework is a good choice. 	<ul style="list-style-type: none"> • They require specific support data for quantification (modeling data i.e. <i>NHEP</i>, PSF multipliers, etc.) • Lacks expressiveness compared to other frameworks • Limited choices limit cross-domain application/adaptation.
Expert-focused	
<ul style="list-style-type: none"> • Their modeling concepts and data used are mostly domain independent. • They work with a lack of pre-existing, or empirical data. • Such methods have also been adapted to other domains. • Most well-known methods provide extensive documentation. • Such methods are moderately difficult to adapt to other domains - only the guidelines need to be changed. • If there is a lack of pre-existing data, experts are the main source and the method's guidelines are detailed enough – this framework is a good choice. 	<ul style="list-style-type: none"> • Reproducibility of analyses using such frameworks has been questioned. • They require high amount of resources (number of experts, training, time etc.) to use. • Compared to multiplier-based methods they are relatively difficult to validate. • Some models still require empirical data to support expert estimation. • Some method's guidelines tend to be highly domain specific.
Probabilistic graphical model-based	
<ul style="list-style-type: none"> • They are relatively easy to use. • They can use data from different sources (empirical, expert judgment, etc.) • They can use preexisting data (from the other two models). • They are relatively easy to adapt across domains. • They are more expressive in terms of modeling capabilities. • If there is mix of expert, empirical and qualitative data available, and the method's guidelines are detailed enough – this framework is a good choice. 	<ul style="list-style-type: none"> • Lack of industrial-scale HRA methodologies which use such a framework. • "...BBNs within HRA have not yet reached a strong level of maturity..." [Mkrtychyan et al., 2015]. • Modeling complexity and subsequent data requirements increase if a large number of variables considered.

Table 4.3 – A comparison of some main HRA methodologies

HRA methodology	Qualitative analysis ¹	Quantification framework ²	Quantitative support data: source ³	Uncertainty quantification: in model / in data (theory elements) ⁴ ♦	Domain of application ⁵	Validation / type in domain ⁶	Recent extensions or adaptations ⁷
THERP [Swain and Guttman, 1983]	Identification guidelines, task and PSF list	Multiplier	Empirical and expert	No / Yes (probability percentile bounds)	Nuclear ♠	Yes / empirical and expert / nuclear	THERP-ACIH hybrid approach [Vanderhaegen et al., 2010]
SLIM [Embrey et al., 1984]	No lists (optional PSF list)	Multiplier (expert estimate focused)	No (guidelines for experts)	No / Yes (probability bounds)	Nuclear ♠	Yes / expert comparative approach / nuclear	HuPeROI [Kyriakidis et al., 2012]
HEART [Williams, 1985]	Generic task and PSF list	Multiplier	Empirical and expert data	No / Yes (probability percentile bounds)	Generic	Yes / empirical and expert comparative approach / nuclear	RARA [Gibson et al., 2013]
CREAM [Hollnagel, 1998]	Generic task and PSF list	Multiplier	Other methods (HEART, THERP, etc.)	No / Yes (probability percentile bounds)	Generic ♠	No / No / No♣ (industrial / nuclear, offshore, space, etc.)	Fuzzy CREAM [Wang et al., 2011], [Marseguerra et al., 2007], Bayesian CREAM [Kim et al., 2006]
MERMOS [Bieder et al., 1998]	Identification guidelines, no lists	Expert estimate focused (conditional probability)	No (guidelines for experts, and empirical)	No / Yes (conditional and expert probability)	Nuclear	Yes / empirical and experts / nuclear	
ATHEANA [Barnes et al., 2000]	Identification guidelines, no lists	Expert estimate focused (conditional probability)	No (guidelines for experts)	No / Yes (probability bounds)	Nuclear	No / No / No♣ (limited industrial usage in nuclear)	
NARA [Kirwan et al., 2004]	Task and PSF list	Multiplier (similar to HEART)	Empirical and expert data	No / Yes (probability percentile bounds)	Nuclear	No / No / No♣ (nuclear)	
SPAR-H [Gertman et al., 2005]	Task and PSF list	Multiplier (similar to THERP)	Other method (THERP)	No / Yes (probability distributions)	Nuclear ♠	No / No / No♣ (nuclear)	Bayesian SPAR-H [Groth and Swiler, 2013]

¹ *Qualitative analysis* states if the method provides *identification guidelines* for guidance on how to perform task analysis or identify more PSF. If it provides *PSF list* and either a domain specific task list or generic task list based on task characteristics (e.g. diagnosis and action).

² *Quantification framework* identifies how the HEP or a similar entity, is computed.

³ *Quantitative support data* refers to the numerical values available with the model guidelines. They are used by the analysts to quantify based on the context to analyze, their *source* is then specified.

⁴ *Uncertainty quantification* is the explicit representation of uncertainty (all types). In *the model*, it is generally epistemic, and *in the data* it is mainly the aleatory uncertainty.

♦ marks, as also remarked in [Chandler et al., 2006a], that none of these methods make a distinction between aleatory or epistemic nature of the uncertainties.

⁵ *Domain of application* is the domain in which the methodology was first proposed.

♠ indicates if the model has been applied in other domains, other than its domain of initial application.

⁶ *Validation* specifies the *type* of validation - if it was done, and the *domain* specific data used to validated the model.

♣ marks the models which have either not been completely validated, or public reports are not available, however are used in the industrial domain as stated in parentheses.

♠ marks the models which have either not completely validated or any such reports are not publicly available, however are used in the industrial domain as stated in parentheses.

⁷ *Recent extensions or adaptations* of the quantitative part model.

-
- HEP integration into system-level analysis: mostly not performed as part of an HRA process – an HEP is input to a preexisting PRA model (an event tree, a fault tree, etc.).
 - Actor: analyst

To give an overview of where such steps fit into the complete HRA methodologies, and what different methods propose, a comparison table is presented in [Table 4.3](#).

Thus, keeping the best practices in a view, and addressing some issues with quantitative HRA in-general, the next section focuses on the needs, previous works and some propositions for rail transportation.

4.3 HRA in rail transportation

In rail transport the way of defining, analyzing and mitigating human error has changed over the years. However, there are a very few complete railway-specific HRA methodologies [[Kyriakidis, 2013](#)]. Some concerned works are discussed below. We also use the notions presented in the previous sections to discuss related propositions in the domain of railway. The context of application for this discussion is limited to *railway operations*, it is defined as: *a train movement from one point to another*.

The study commissioned by the European Union Agency for Railways (EUAR) [[Kecklund et al., 2013](#)] presents a survey on Railway undertakings, Infrastructure managers and national safety authorities. It concluded that, "*... even though the respondents participating in the survey performed some types of risk assessments as related to human interaction, they did not necessarily use any established human factors technique for this ...*" and further "*Most of the responding RUs (Railway Undertakings) and IMs (Infrastructures Managers) do not use any specific human factors techniques*". Further, in this survey, a question asking about the usage of a specific HRA or similar technique was posed. The frequency for the responses is presented in [Figure 4.11](#). It shows that most of them do not use any specific human factors techniques, but state that human and organizational errors are handled within the general risk assessment technique, often or as part of every assessment. Thus, there is a need for more work in this domain, and before that the existing works need to be discussed.

Q5. Looking at the performed risk assessments within the past two years (2010 and 2011) in your organisation, can you please state in which areas risk assessments were made?	never	seldom	occasionally	often	every assessment
a. Human and organisational errors are handled within the general risk assessment techniques	4	0	2	9	6
b. Technique for Human Error Rate Prediction (THERP)	21	0	0	0	0
c. Job safety analysis	15	0	1	2	3
d. Systematic Human Error Reduction and Prediction (SHERPA)	21	0	0	0	0
e. Expert Judgment Methods	16	2	1	1	1
f. Technique for Retrospective Analysis of Cognitive Errors (TRACEr)	20	0	0	1	0
g. Human Error Assessment and Reduction Technique (HEART)	20	1	0	0	0
h. A Technique for Human Error Analysis (ATHEANA)	21	0	0	0	0
i. Cognitive Reliability Error Analysis Method (CREAM)	20	0	1	0	0
j. Human Error HAZard and OPerability study (HE-HAZOP)	20	0	0	1	0
k. Predictive Human Error Analysis (PHEA)	21	0	0	0	0
l. Cognitive Task Analysis (CTA)	20	0	1	0	0
16.m. Hierarchical Task Analysis (HTA)	20	0	1	0	0
16.u. Other	17	0	1	1	0

Figure 4.11 – Number of responses from railway entities on "how often they use specific human factors techniques?" taken from [Kecklund et al., 2013]

4.3.1 Variables of a quantitative framework for railway

As discussed previously in section 4.2.1 HFE, and PSFs identification is an imperative prerequisite for a complete HRA methodology. Although, there are some works which treat similar issues for the railway domain, they are not necessarily done from an HRA perspective. Such works can be used as building blocks of an rail-HRA. The following discussion aims to present some relevant works in the railway domain. The objective is to identify works which can support a quantitative HRA for railway.

4.3.1.1 Human failure event for rail operation

HFEs are traditionally identified as part of PRA activities, as required by the regulations. For railway applications an equivalent is identification from functional (FTA, event tree, etc.) analysis on the technical systems section 3.4.2. The basic events (BE) which involve a human action, as identified in fault trees can be used. These events should be of significant importance to require a further analysis, such as human actions which are classified as safety critical.

For example, in [UNISIG (Union of Signalling Industry), 2016] a fault tree and subsequent functional analysis of ETCS application level 1 and 2 (overview given in section 3.4.1) are performed by UNISIG (Union industry of signaling). For example they analyze one of the main functions of the ETCS system in an operational environment, this function is defined as: "To provide the driver with information to drive the train safely and to enforce respect of this information to the extent advised to ETCS." For this definition they defined the conceptual fault tree [UNISIG (Union of Signalling Industry), 2016], as shown Figure 4.12. This

fault tree in particular analyzes the accomplishment of the ETCS function (top level event) *preventing train over-speed* (vs. speed limits computed by the ETCS system). In [Figure 4.12](#) on the right-most bottom event a driver error is identified, where he/she exceeds the speed limits. Such an identification is carried out for an operational railway environment, in terms of system-level functions. By braking down the scenarios are developed enough to qualify for HFE identification. Further, for such safety events a FMEA is carried out, the FMEA for this driver error is given in the bottom part of [Figure 4.12](#). Such an analysis provides important details on a HFE – a driver error: DRV-1.

Nevertheless, as stated in the study: “*this fault tree does not imply or mandate a specific system implementation. . .*”. The definitions and functional analysis are not strict implementations and can change. Furthermore, it lacks two crucial components: harmonized application, and national signaling and operating rules (the procedures). These are out of the scope of an ETCS specification which is analyzed in this work as analyzed in [[UNISIG \(Union of Signalling Industry\), 2016](#)]. As shown in [Figure 4.12](#) that DRV-1 is exported condition to the *national procedures*. More such basic events are analyzed in [[UNISIG \(Union of Signalling Industry\), 2016](#)] and we see other exported conditions such as: *Needs to be covered by national procedures, Data entry procedure should protect against basic human error; Driver vigilance is presumed.* etc. The operating rules are generally defined by the operational authorities (railway undertakings and/or safety authorities). Thus this aspect is crucial to ensure that the HFEs are well defined and precise, and to carry out further analysis (context). Nevertheless, such a work can be used to complement HFE identification. Some rail-specific retrospective analysis approaches exist, as analyzed in [[Baysari et al., 2011](#)]. However, their objective is more towards prevention and/or mitigation strategies rather than a predictive quantitative focus. They also concluded that the task of error identification needs local (national) considerations and appropriate context relevant definition of terms in order to be usable. Thus, a formal system-level functional analysis together with human components, and operational rules are needed for identification of HFEs.

There are some other alternatives, in [[Boring, 2015](#)]. The author describes a way to identify HFEs from a Human Factors study. Similar approach can be followed for railway applications, where such human factors studies are relatively easily available (e.g. [[Pickup et al., 2013](#)], [[Vanderhaegen, 2001](#)], etc.). They, however, cannot be the sole source, mainly because identifying failure events is not their primary objective. Furthermore, such studies tend to be very generic, making it difficult to focus on specific human actors, and application contexts involved and the signaling technology used. For example, ACIH [[Vanderhaegen, 2001](#)], can be

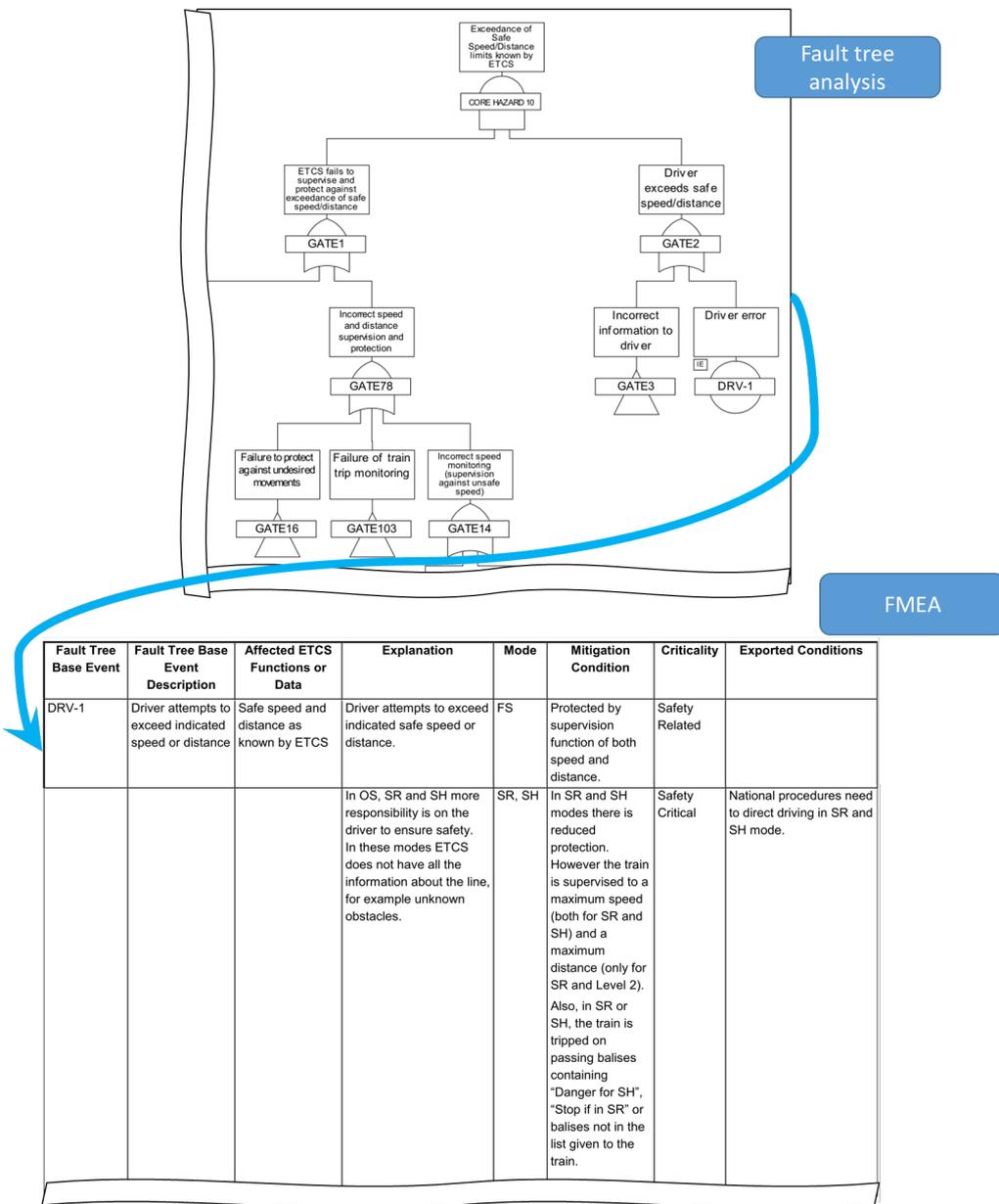


Figure 4.12 – Conceptual Fault Tree for the functional analysis of the ETCS (application level 2) within an operational railway environment [UNISIG (Union of Signalling Industry), 2016] and the FMEA for a *driver error* base event.

used to do a task-analysis to obtain details on the HFEs (or more granular actions). The author proposes a functional decomposition of railway system, to identify the task to be performed and associated behavior (an HFE's task characteristics). The work also characterizes such HFEs – actions, omissions, cognitive tasks, etc. This activity can be used to help experts to assign quantitative values, similar to *NHEP* values. This activity will depend on the general framework of a quantitative HRA method, nevertheless, such a works provides useful qualitative framework.

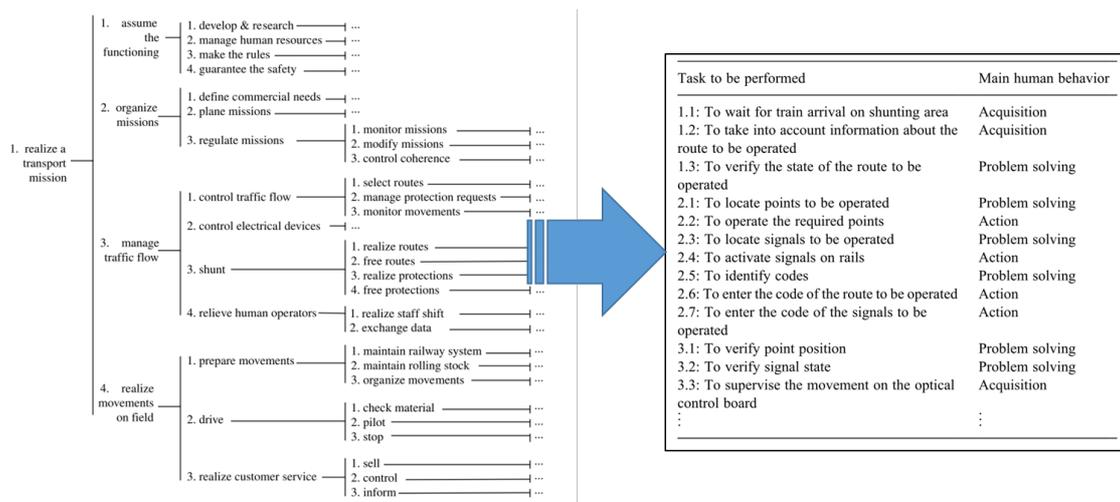


Figure 4.13 – Functional decomposition of railway system, to identify the task to be performed and associated behavior (an HFE's task characteristics), adapted from [Vanderhaegen, 2001]

4.3.1.2 PSFs for railway and related works

Ergonomics and human factors are often used interchangeably in the railway domain and have attracted large part of the research. A review study of these factors [Wilson and Norris, 2005] indicates the increasing interest of railway stakeholders in understanding human factors.

A qualitative HRA method, for rail application is presented in [Schwencke et al., 2012]. They discuss PSFs and their importance to characterize a context. Their emphasis is on human resilience, in turn systems to cope with unfamiliar situations and disturbances. They focused on the importance of PSFs towards proposing a context-related HRA model for rail systems. One of the important conclusions of the study is that the differences among national level rules, specialty in railways, make it difficult to find globally applicable results.

In [Hammerl and Vanderhaegen, 2012] a qualitative analysis approach has been proposed to account for the certification requirements. They aim to provide an overview of human factors to a railway engineer dealing with certification

requirements. Furthermore, they argue for the applicability of a PSF-based model in safety analysis of a human-barrier interaction.

In 2013 an in-depth study of human factors and their integration European railways was commissioned by EUAR [Pickup et al., 2013]. It aimed to provide a human-centered perspective towards systematic integration of human factors at multiple levels e.g. design, development, operation and maintenance of railways in Europe. The final report consists of relatively extensive data analysis (over 16 countries). And a wide range of actors/operator job roles. Subsequently an expert-opinion based analysis of safety relevant activities of humans involved in railway operations was performed on the raw data. Their results "*provided a generic and high level view of human functions and identify safety relevant human activities associated with these functions.*" The amalgamation of system-level functions, human functions, safety relevant makes such a study a good candidate for a HFE/PSF identification. The objectives are still not close to an HRA-like application, but its exhaustiveness makes-up for this.

A PSF list for railways. As discussed in [section 4.2.1.2](#), most HRA methods include a PSF list with the model. Thus, a generic concise PSF-list for the rail operations can be proposed. The question is what PSFs should be included in such a list?

The first view is that for ensuring safety, the PSFs that have been implicated most frequently in past accidents/incidents/mishaps should be used [Kyriakidis et al., 2015a].

That is, PSFs from accident analysis data. In [Kyriakidis et al., 2011] the authors conclude that 18 PSFs were responsible for more than 80% of the railway accidents analyzed. Later in [Kyriakidis et al., 2015a] the authors reached a similar conclusion, 12 PSFs alone or a combination thereof were responsible for 90% of the accidents analyzed. This was a smaller list obtained from their original list of 43 PSFs. A railway specific PSFs taxonomy called "*R-PSF lite*" was proposed in [Kyriakidis et al., 2015a]. It was developed from human factor literature review, railway accident and incident reports and validated by expert opinions. Their objectives were oriented towards constructing a domain specific PSF list with subsequent expert and accident analysis based validation of the said PSF set. The R-PSFs lite is given as follows [Kyriakidis et al., 2015a]:

- Safety culture
- System design
- Fatigue - shift pattern - fit to work

-
- Communication - teamwork
 - Distraction - loss of concentration vigilance - situational awareness
 - Quality of procedures
 - Perception - interpretation
 - Training - experience
 - Expectation - familiarity - routine
 - Quality of information
 - Supervision
 - Workload - time pressure - stress

The second point-of-view is a prospective human factors functional analysis approach, by human factors experts. The EUAR's HF study is another good candidate for identifying PSFs that are actually important from a rail operations perspective. The spreadsheet along with the report [Pickup et al., 2013] presents the results of the analysis. The results are presented in a multilevel hierarchy, starting from system level operational goals down to a human's safety relevant activities. A brief discussion is presented here, the text and nomenclature extracted from [Pickup et al., 2013] is given in *italics*. The top level system operational goals are defined as *Purpose/Goals - aim of the socio-technical system and a focus for human efforts*. In total, seven such high level goals were identified: maintain safety, provide efficient train service, optimize passenger comfort and journey, minimize environmental impact etc. These goals are then attributed at the second level to *human functional goals*, for example for all of the aforementioned Purposes/Goals it is necessary that *train movement must be controlled in all operational circumstances*. Subsequently 8 human functional goals were identified each branching from one or more purposes/goals. Each *human functional goal* is further broken-down into multiple lower level *human functions*. A spreadsheet is produced describing the *human function* under analysis, the context under which it is executed, and the analysis of *safety relevant activities* associated with it. This analysis includes data on: *safety relevant actions or activities, potential for errors, recovery, mitigation strategies, with discussion on respective conditions or casual factors (PSFs) for each. Potential for errors* describes the factors and scenarios which might provoke an error. The factors appearing in the safety analysis of a single human function with a negative connotation (i.e. increasing the potential for occurrence of an error implicated in a system level safety objective), can be considered to have a considerable effect on system/operational safety. This allows the identification of safety critical PSFs for

the said human function. More specifically, every human function has two columns *safety relevant activities* and *analysis of safety relevant activities*. Further, in some cases *potential for recovery* can also be used to gather more information on the PSFs. As shown in Figure 4.14 the boxes marked in a continuous black border are goals and functions which are involved in maintaining system safety. The parts which can be used to extract PSFs, as explained here, are marked in red.

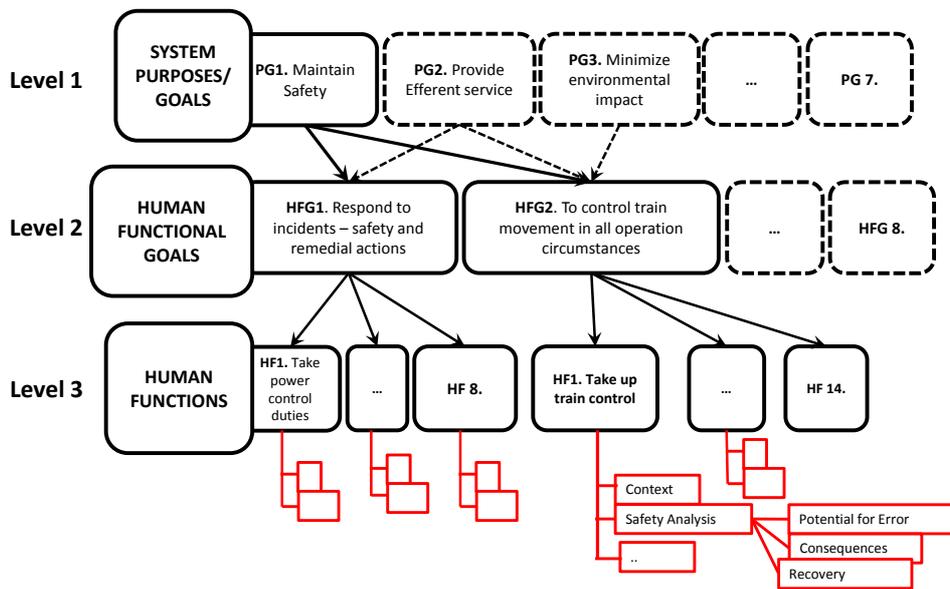


Figure 4.14 – ERA study data organization and exaction of PSF. The parts of the study which can be used for the PSF list are colored in red.

To give an example of the generation of PSFs from human functions and subsequent description thereof, a simple case is explained, also shown in red outline in Figure 4.14. The first entry in the spreadsheet under *human functional goal* we consider the lower level human function *Take up control of train movement duties*. The last column in the spreadsheet, the potential for errors associated with this activity states: "A lack of understanding of the information that is needed to appreciate the status of the system, possibly linked to inattention, memory failure." Distraction/Concentration or the absence of attention is a well-defined PSF, and also figures in R-PSF. Thus, it is identified as a PSF that we should consider. The identification of underlying factor for *memory failure* cannot be identified and more information under mitigation is referred which states "improving experience" along with *Protocols for communications and procedures for handovers*. Furthermore there

is an absence of PSFs directly describing memory failure. Hence, in this case only one PSF is considered. This way, a PSF list identified from a analysis of human functions, which are implicated in ensuring system level safety a can be used to build a PSF reference list for HRA activities.

However, the data, by their own admission [Pickup et al., 2013] is heavily influenced by UK-based sources; and there is need to distinguish degraded cases, and normal operations, since they can change the context for a human significantly. So far we have considered factors from a retrospective analysis (accident data) and prospective human factors analysis.

The third view is to consider general human reliability aspects, i.e. PSFs from other HRA models and taxonomies. The work on analyzing and aggregating PSFs for HRA purposes in [Groth and Mosleh, 2012a] provides a fairly exhaustive dataset on PSFs, and cross-domain definitions of PSFs to a certain extent. It offers a relatively exhaustive set of PSFs, with their definitions. With the level of detail and exhaustiveness, this work can be considered for generic and non-domain specific context of application.

Further, towards quantitative considerations the size of a PSF list can need to be considered. A good practice guide for the nuclear domain [Kolaczowski et al., 2005] gives a list of 14 PSFs to consider. Further, in [Mosleh and Chang, 2004] (also discussed in section 4.2) states that the PSFs should be measurable. Factors such as Supervision, Safety culture, are difficult to accurately measure, or ask experts to quantify.

Thus, the proposed approach in this work is to have a relatively short PSF list, akin to SPAR-H [Gertman et al., 2005], which is simpler for the analyst and the experts, to use [Whaley et al., 2011]. Hence, once a model which provides limited yet reliable quantification results is obtained, more factors can be added to increase the scope and applicability of the model. To further limit the scope and to ease the usage, one *human functional goal* was chosen from EUAR HF study (red outline in Figure 4.14) to adhere to our consideration of the train driver *To control train movements in all operational circumstances*, which includes nominal and degraded cases.

The general definition of the PSFs were taken from previously discussed works [Groth and Mosleh, 2012a] and [Gertman et al., 2005], and some rail-specific considerations from the EUAR study. The quantification levels were taken from [Forester et al., 2004]. The final PSF list with the definitions and levels is given Table 4.4. The quantification levels are focused presently for expert elicitation [Forester et al., 2004], but can be easily made PSF specif, similar to some other methods (section 4.2.1.2).

Table 4.4 – PSF list with considered definitions and quantification levels, adapted from [Rangra et al., 2015a]

Performance Shaping Factor	Definition	Qualitative levels
Training	Have the correct knowledge to perform a job successfully and safely. Training might be needed to ensure skills are up to date and relevant, i.e., new procedures, different signaling systems, etc.	Good, Nominal, Poor, Insufficient Information
Experience	The accumulation of information and knowledge gained through interactions with the system and time spent in the work environment, this can be in same conditions (or same route). Aspects like, bad habits learned, etc. should also be considered in addition to the, positive aspects.	Good, Nominal, Poor, Insufficient Information
Communication	The ability of team members to pass information to each other and a shared understanding of the situation using, system status, read-outs, etc. e.g. misunderstanding, omission of, information, mistaken location, incorrect communication actions. Human-machine communication aspects are not included in this PSF.	Good, Nominal, Poor, Insufficient Information
Situational Awareness	The perception of the elements in the environment within a volume of time and space. The comprehension of their meaning and projection of their status in the near future.	Good, Nominal, Poor, Insufficient Information
Task Load (Workload)	The actual task demand assigned to a person in terms of the number and type of tasks (varying complexity, importance, etc.). Task load can also be impacted by unplanned or emergency events.	Good, Nominal, Poor, Insufficient Information
Time load (Workload)	Time required or allocated for one or multiple tasks; this time perception can affect worker stress beyond the stress of having too many tasks. Available time to complete a task particularly in the context of driving activities related to high speed trains (both detection and completion of the task).	Good, Nominal, Poor, Insufficient Information
HSI quality	An umbrella term to consider the quality of human system interface. The broad context here includes the procedures, appropriate information displayed to the human at appropriate time or in an adequate way. It includes most ‘Machine-based factors’ directly influencing human behavior.	Good, Nominal, Poor, Insufficient Information

A complete PSF list should be a union of these three point-of-views, namely historical data, human factors experts' analysis and best practices of cross-domain HRA methods. Ensuring that the analyst has a relatively extensive list of PSFs, which can be used as a reference list of factors.

Usage of such a PSF list. Such a list is only an additional aid to the analyst. The usage of such a list might be modified based on the application. In the traditional HRA approach, once the HFEs are identified, information concerning the context and operational environment (section 4.2.1.2) is needed to identify PSFs. PSFs that are implicated in an HFE's context, towards degrading human performance are to be considered. Similar to the idea behind SPAR-H's pre-screening section 4.2.3.1.

Related ERA HF and similar studies (safety perspective of human functions/-tasks/goals) can be used identifying specific PSFs from such contexts. For this transition, a mapping, which functionally matches the HFE to a relevant human function is needed. For example, for the basic event as identified in Figure 4.12, an HFE. A *human function* needs to be identified (say DRV-1 in Figure 4.14), which matches the definition and context of the HFE. Non-accomplishment of this function then represents the HFE in it's context. And the safety analysis gives the PSFs implicated to impact human performance positively or negatively, towards this objective. A more application oriented case can be tasks/procedures extracted from official procedures, such as [SNCF Réseau, 2016] and then matched to the human function in ERA HF study.

Most human factor studies closer to PSF point of view, point out the need of inclusion of national rules – or more generally the operational context and environment (signaling systems, etc.) and procedures (operating rules). Such aspects are to some extent dependent on application, but might limit the validity of an HRA proposed for a different country's regulation. This is also more important in the context of ERTMS, and the push to unify rail signaling in Europe. Thus, not only the operating rules need to be adapted, but also the risk analysis needs a normalized context to be applicable from country-to-country.

4.3.2 Some quantitative considerations of human errors and frameworks in railway

The way of defining, analyzing and mitigating human error has changed over the years. However as discussed in [Kyriakidis, 2013] there are very few complete HRA methodologies for railway. Some concerned works are discussed below.

The usage of a predictive analysis is motivating for an application when seen in the context of Reliability, Availability, Maintainability, and Safety (RAMS) assessment including traditional risk analysis techniques. This may aid the system and procedure design as demonstrated in [Connelly et al., 2012] or demonstrate conformity with safety standards. However, such an application needs well established and validated methods. From the regulatory framework of Risk Acceptance Criteria, there are different approaches possible to integrate HEP (or in general human error data), or for assessments of the associated human error risk. One of the approach which deals with explicit numbers as stated in [Mowitz and Kecklund, 2013] is: “. . . human reliability data integrated within any other assessment technique”. If an HRA method allows obtaining such data, we can integrate into a system-level assessment, similar to the PRA–ATHEANA relation we saw in section 4.2.3.2. However, as discussed previously, the survey of the industrial actors in [Kecklund et al., 2013] shows that very few HRA methods (or a similar approach) are used in the industry.

Type of human behaviour while performing the activity:	Favourable environmental conditions when performing the activity:			Unfavourable environmental conditions when performing the activity:		
	Stress through a too low demand	Optimal stress level	Stress through an excessive demand	Stress through a too low demand	Optimal stress level	Stress through an excessive demand
Skill-based behaviour	$2 \cdot 10^{-3}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$1 \cdot 10^{-2}$	$5 \cdot 10^{-3}$	$1 \cdot 10^{-2}$
Rule-based behaviour	$2 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$1 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$1 \cdot 10^{-1}$
Knowledge-based behaviour	$2 \cdot 10^{-1}$	$1 \cdot 10^{-1}$	$2 \cdot 10^{-1}$	1	$5 \cdot 10^{-1}$	1

Source: Hinzen: „Der Einfluss des menschlichen Fehlers auf die Sicherheit der Eisenbahn“, RWTH Aachen, PhD, 1993

Figure 4.15 – Predefined values for human error (HEP), as used in some rail applications

In some cases error probability of every human action is assigned a fixed value of 10^{-3} [Schwencke et al., 2012]. Further, this is on the number of events, rather than time based, i.e. one error per 10^3 events. It differs than the general time-based reliability scale of technical components. Here, predefined values are used, Figure 4.15 shows these values. Although these values can be seen as a smaller version of the table of *NHEP* values as seen in section 4.2.3.1 *Multiplier-like*. Among other critiques, the usage of pre-defined numbers, at the very least oversimplifies the high variability of human performance. Furthermore, these probability values are not proven to be valid, and sufficient accident data is said to be unavailable to validate them [Feldmann et al., 2008].

Some works which propose or use, a quantitative HRA or similar approach are discussed here. They are briefly discussed and place in the context of a complete HRA process.

An methodology proposed in [Connelly et al., 2012] applies a human reliability analysis model to the PTC DMI (Positive Train Control Driver machine interface:

similar to a ERTMS/ETCS); this system possesses similarities to the ERTMS on-board system DMI. Such an application approach can help identifications of case studies for industrial feasibility demonstrations.

In [McLeod et al., 2007] risk of human interaction with Automatic warning systems (AWS) was evaluated. They proposed an influence model having an explicit representation of the way the factors affect driver reliability. The situational factors and risks had a one to one relation and the identification of factors represented an increased risk. Different scenarios involving AWS were characterized in terms of risk using these situational factors.

Railway Action Reliability Assessment (RARA) [Gibson et al., 2013] is presented as a technique for quantifying human reliability. It aims to support the quantification of human performance as part of human reliability and risk assessment processes. It is a GTT-based approach, a nominal HEP is assigned to a task based on the type of generic task. Task situation-related PSFs are selected, a weight and influence metric, (similar to HEART, [section 4.2.3.1 Multiplier-like](#)) assigns the influences of the selected PSFs to get final HEP. The technique is said to be particularly suitable for risk or safety decision making in cases where data (e.g. from real accidents or simulator experiments) is not available.

A report [Thommesen and Andersen, 2012] describes a HRA for "six generic tasks and four PSFs targeted at railway operations commissioned by Banedanmark (Danish national Infrastructure Manager)." This study aimed to propose a HEART-like method for railway application. They aimed to address some critics of using HEART estimates for railway application, such as: the definitions of GTTs, the HEART "... estimates may be too conservative for railway driving tasks." and assess their validity for railway. They perform an analysis of different HRA methods and a generic database to make some recommendations. Notably, they recommend the quantitative values for: *NHEP* estimates and the multiplier values for PSF.

A study with similar objectives [Kyriakidis et al., 2012] proposed an HRA model called Human Performance Railway Operational Index (HuPeROI). The human performance is measured by human reliability that is Human Error Probability (HEP). It is based on the R-PSF taxonomy, previously discussed. The formula for determining the "HuPeROI success index" is given as:

$HuPeROI_j = \sum_{i=1}^n w_i \times r_{ij}$, where, w_i is the weighting for the i th R-PSF and r_{ij} the rating of task j on the i th R-PSF. In their approach expert data is needed for PSF measurement, the weighing factors for quantification. Expert opinion is needed for the quantification of PSF, its relative impact on human error (weight) and correlations between PSFs. Note that, for most expert-based methods, experts are elicited different values (HEP, probability distributions), and often conditional data,

as discussed in [section 4.2.3.2 Expert focused](#). Further, this expert data collection process and how such a quantitative framework can be used is not presented. It is said to be based on the concept of SLIM reflects the overall belief of the SMEs, regarding the positive or negative effects of the R-PSFs on the likelihood of success for the task under consideration. Their final objective is to "estimate the relative likelihood of human error for several operational scenarios." It lacks a component of HEP quantification.

Thus, building blocks of a complete quantitative HRA method can be found in the works over the years. Nevertheless, there remain some challenges, with HRA in-general, and some for the railway applicability. Towards a robust railway HRA method, a discussion is presented next, which aims to complete this state of the art.

4.4 Discussion towards a complete railway HRA methodology

We refer to [Table 4.3](#) to guide the discussion towards the needs of a rail HRA methodology.

Guidelines on identification of PSF (rail-specific) and HFE: most complete HRA methods provide this step, but are domain-specific in nature, *column 2, table 4.3*. In [[Le Bot, 2004](#)] the authors argue that while including PSFs in the model, the focus should not completely be on HEPs, or errors in general. Specific situational elements (e.g. PSF-like elements of context) that may contribute to a failure should be analyzed. This shifts focus of the values of HEPs themselves to the factors around it, making HEP a local (specific to context) rather than a global (at all times and at all situations) phenomenon. This, shows a difference from the definition of technical component failure, which is the case with failure rates. Further, HEP becomes an indication of a situation needing special attention, actual value thereof being less important. It indicates an operation, a task, an event, which is not faulty as such, but inappropriate considering the particular context. Thus, PSFs and their characterization becomes all the more important. Several PSF taxonomies are addressed in the literature including with HRA models. For the well known models, their PSFs lists have been gone through multiple revisions and critiques giving them a refined definitions and hierarchical structuring, among other classifications. The nature of PSFs make them relatively easy to extract from accident analysis reports, which are one of the main sources of information of human erroneous behavior. Also, a functional analysis of humans activities can lead to similar lists of safety critical PSFs.

PSFs do not affect humans equally, and such data can be subjective, i.e. the multipliers in [section 4.2.3.1](#) or the CPT in [section 4.2.3.3](#). Verifying the validity of PSFs is a matter of concern, as stated in [[Boring et al., 2007](#)]. "*The analyst should ensure, informally or through formal structure analysis techniques, that the PSF measures what it purports to measure*". Secondly, inter-PSF relations are not easy to identify and model, and subsequently pose problems in the quantification model, such as double counting among others. Quantification of PSFs themselves poses a challenge, as on one hand it is difficult to accurately measure such subjective factors on human performance and on the other we do not have concrete transportation specific studies and extensive expert opinions to provide guidelines. Thus, a flexible yet expressive mathematical framework should be used, which can express such causal relations between variables. A probabilistic graphical model such a BNN offers such a framework.

Usage of expert data in quantitative HRA modeling: most quantitative HRA methods use some form of expert data. However, it is not necessarily conditional, or lacks an explicit consideration of conflict or uncertainty in model *column 3, table 4.3*. Further, when expert data is employed considerations on combination of expert data also need attention.

Usually the expert elicitation is termed as a subjective judgment and represented as a subjective probability density function (PDF) reflecting the experts belief. Since probabilistic elicitation and by extension PDFs remains easier to elicit and straightforward to use it is rather frequently preferred. An extension proposed in [[Podofillini and Dang, 2013](#)] aims to allow formal aggregation of expert estimates, to account for expert variability and inherent variability in HEP estimates. A Bayesian approach is employed to update the quantification model, as and when information is received. The second aspect is uncertainty representation. There are various ways to represent both types of uncertainties imprecise (interval) probability, Possibility theory, Belief function theory are some of them. However as concluded in [[Aven, 2011](#)] only the probability bound approaches provide easy interpretation in a practical decision-making context. Such aspects can be managed with the use of a probabilistic graphical model as presented in [section 4.2.3.3](#). Extensive discussions exists in some complete HRA methodology's application guidelines [[Gertman et al., 2005](#)] [[J. Forester et al., 2007](#)]. The use of such frameworks can possibly make such considerations more accessible to experts and analysts.

In the absence of empirical data for the rail domain: where tabular data is not present, expert data can be used, however, in almost all uncertainties are not considered explicitly in the model *column 4, column 5, table 4.3*. Uncertainty considerations have been a part of the PRA process in the nuclear industry since

quite some time. In [Forester et al., 2004] the general elicitation process is renamed 'quantification-including-uncertainty' to emphasize the importance thereof, their statement "*quantification includes uncertainty, because anything else would be incomplete*". However in their case explicit considerations of epistemic uncertainty are left to the experts rather than the formulas. For rare-event quantification the considerations of uncertainty and the nature thereof are important questions. The identification of uncertainty holds an important place, and second the ability to work with less data. As frequently characterized by its source, there are two types of uncertainty in data – the one originating from natural randomness called *aleatory* and the one from a lack of information is termed *epistemic*. Further, as described in [Parry, 1996] for the context of PRA *uncertainty is that associated with the analyst's confidence in the predictions of the PRA model itself, and is a reflection of his assessment of how well his model represents the system he is modeling*. As evident it reduced by improving the model of the system under analysis. Aleatory on the other hand is independent of the analyst's (or experts) knowledge of the system. And therein lies the interest in making this classification, it helps in understanding what is reducible and what is not. Unfortunately for human reliability analysis or rare-events this problem gets further complicated and adequate theoretical representation is therefore needed.

In [Swain and Guttman, 1983] a discussion on the source of uncertainty in HRA, mainly epistemic uncertainty, is presented. We list some points as follows:

- Dearth of the type of human performance data useful to PRA/HRA
- Inexactness of models of human performance
- Inadequate identification of PSFs and their interactions and effects

Thus, the choice of a modeling framework which can provide means to explicitly represent and manage uncertainty in the model – such as a probabilistic graphical model should be employed. Finally, since most methods are proposed for the nuclear domain, very few are generic, *column 6, table 4.3*; railway specific considerations should be done, in addition to the PSF list proposed previously.

4.5 Conclusions

This chapter presents some considerations towards a quantitative human reliability analysis model for rail transportation. Further, arguments for a PSF-based HRA model for transportation have been presented. Thus, aiming for a quantitative HRA is a challenging endeavor, a pragmatic approach nonetheless.

Towards a railway HRA model, this chapter also introduces a rail-specific PSF. Data on human functions and safety relevant activities thereof have been analyzed. The generated PSF list is defined and adapted for railway application needs, referring to domain specific studies. Existing problems, the data sources and works needed to address those problems are identified. Finally, goals for next steps to arrive at a limited scope but robust quantitative HRA methodology are charted in this work. The need for a method able to measure HEP values with a reasonable degree of uncertainty for the factors frequently observed is discussed. The contents towards such a systematic framework capable of analyzing human errors quantitatively are also presented in this chapter.

Thus, towards a complete HRA methodology for railway the next chapter will start with a qualitative analysis which allows identification of PSFs and HFEs. It will employ the PSF list proposed here in [section 4.3.1.2](#) which includes PSFs and is adapted to the domain-needs is required to guide the analyst. We also aim to employ a probabilistic graphical model, since it seems to be one way forward for quantitative HRA. Such a framework allows the integration of traditional HRA concepts, adequate representation and management of uncertainty, combination of different sources of modeling data, causal and subjective data from the research domain into more application-oriented and usable format.

PRELUDE: Performance shaping factor based human reliability assessment using valuation-based systems

Contents

5.1 The PRELUDE methodology	100
5.1.1 Qualitative part	101
5.1.1.1 Performance Shaping Factor list and evaluation	102
5.1.1.2 Identification of HFEs and safety critical context	103
5.1.2 Quantitative part	106
5.1.2.1 The expert elicitation process	107
5.1.2.2 Combination of expert data	109
5.1.2.3 Transformation	112
5.1.3 Quantification and sensitivity analysis	116
5.1.3.1 Assigning the direct evidence and quantification	116
5.1.3.2 Sensitivity analysis	117
5.2 Case study	121
5.2.1 Step 1. Qualitative part: HFE and PSF(s) identification	121
5.2.2 Step 2. Quantitative part: Expert elicitation, data combination and transformation	125
5.2.3 Step 3. Quantification data and results	128
5.3 Conclusions	133

This section proposes a new and comprehensive HRA methodology titled ‘PRELUDE’, an acronym for (Performance shaping factor based human REliability assessment using vaLUation-baseD systEms). This section presents the methodology and cites the theoretical concepts in the previous sections. A stand-alone, however less detailed and explanatory form of the PRELUDE methodology was published in the paper [Rangra et al., 2017a].

The original contributions and some key points of the PRELUDE methodology are as follows:

- Guidelines on identification of PSF (rail-specific) and HFE, from human functions and accident analysis reports (most complete HRA methods provide this step, but are domain-specific in nature, *column 2, table 4.3*).
- The expert data combination approach provides guidelines on using different mathematical data combination rules. A particular focus is made to manage conflicting opinions, implicitly and explicitly. Most expert data-based methods lack such considerations. *column 3, table 4.3*.
- In the absence of data for the rail domain, a formal expert data combination and transformation approach, is proposed. It takes as input conditional expert data and transforms it into valuations for a VBS model. Most other methods use expert elicitation, however, it is not necessarily formally modeled, *column 3, table 4.3*).
- The VBS/BFT framework allows for an explicit representation of imprecision of data in modeling, and quantification as imprecise probability intervals. This allows us to make a distinction between aleatory and epistemic nature of the uncertainties. Such considerations are often not made in most other quantitative approaches, *column 5, table 4.3*).
- A railway specific-application is demonstrated, Application on a real, recent high-speed railway accident scenario (most methods are either proposed for the nuclear domain, very few are generic, *column 6, table 4.3*).

5.1 The PRELUDE methodology

PRELUDE an acronym for *Performance shaping factor based human Reliability assEssment using vaLUation-baseD systEms*, is a human reliability analysis methodology. The complete methodology entails a qualitative part which accounts for human factors and domain specific considerations, a quantitative part which builds

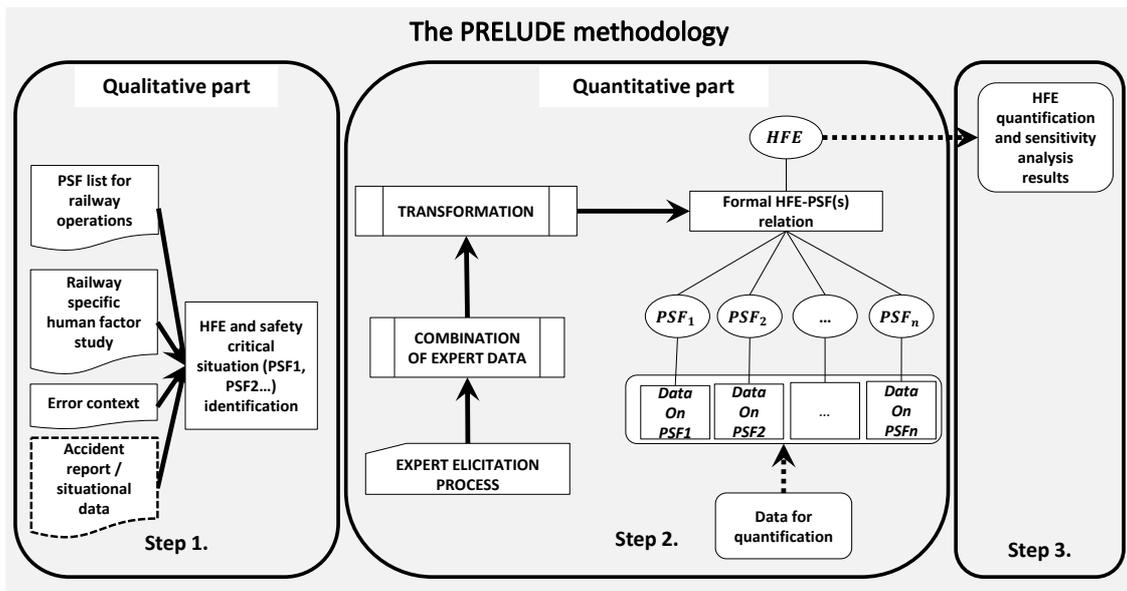


Figure 5.1 – Overview of the PRELUDE methodology. Step 1. is the qualitative part which aims to identify and characterize a safety critical context, as HFE and a set of PSFs; Step 2. is the quantitative part, which builds the VBS model from expert data; and Step 3. presents the final HFE quantification and sensitivity analysis results.

an expert system formalizing expert knowledge and providing formal decision-making. To illustrate the proposition, an overview of the PRELUDE methodology is presented in [Figure 5.1](#). The qualitative and quantitative propositions are detailed in [section 5.1.1](#), [section 5.1.2](#) respectively. Finally quantification using an example of the obtained VBS model is discussed in [item 5.1.3.2](#). PRELUDE methodology, as presented in this manuscript is applied to the railway domain, but it is also adaptable to other applications. For a more detailed, user-oriented illustration of the PRELUDE methodology is presented as a flowchart in the appendix [section A.1.1](#).

5.1.1 Qualitative part

In some application domains (like railway) a PRA or PRA-like analysis with explicit considerations of HFEs are unavailable. Thus, adequate propositions must be made to extract HFEs and identify safety critical contexts of situations in operational conditions towards HRA considerations. The qualitative proposition of PRELUDE is centered on HFEs and PSFs, it aims for a characterization of a safety critical context of an HFE as a (sub)set of PSFs. Appropriate rail-specific sources are employed towards concertizing the human factors background of the methodology.

The variable of interest for PRELUDE as with most HRA models is an HFE. HFEs are predefined in terms of disruptions to component, or system functioning, in which humans are involved, either by causing the failure or not preventing or

mitigating the failure, and represents the basic unit of analysis in the HRA. It should be remarked that a noticeable difference must be maintained when performing prospective vs. retrospective analysis, since the objective of the analysis changes the variables and their interpretations. The case study in this work presents an retrospective analysis approach. However, for a prospective analysis, a discussion to identify HFEs from a human factors study is also presented.

For each HFE identified, a safety critical context using PSFs, i.e. variables of the proposed HRA model. In order to represent the domain specific human factor concerns, present work uses a study by the EUAR's human factors network – 'Support Study for Human Factors Integration in European railways' [Pickup et al., 2013] (hereafter referred to as *EUAR HF study*). It presents a detailed analysis of human functions and goals in railway operations in terms of operational safety and other system-level objectives. Also, as of the most recent information (2015) from the authors its validation is in progress. The first sub-section presents a generic PSF list for rail operations. In the second sub-section main focus shall be to identify PSFs to characterize safety critical contexts in operational conditions.

5.1.1.1 Performance Shaping Factor list and evaluation

Generally speaking, PSFs in an HRA should be easy to use and adapted to respond to the needs of the application domain. Current work uses a rail-specific PSF list as originally proposed in [Rangra et al., 2015a]. A slightly modified version used in present work is given in Table 5.1. Each PSF is accompanied by a definition and a finite number of levels, also known as factor ratings [Podofillini and Dang, 2013] or rating scales or qualitative quality descriptors [Spurgin, 2009]. The term PSF levels or simply levels are used hereafter. The levels considered in the present work are similar to what is normally seen in other HRA models [Gertman et al., 2005] and activities. This qualitative work considers three levels (*good*, *nominal*, *poor* for each of the PSFs (ref. Table 5.1). They are defined (adapted from [Whaley et al., 2011]) as follows:

- *Good*: A PSF assigned this level is conducive to good performance, such that it reduces the opportunities for error, and thus, does not pose any safety issues.
- *Nominal*: It is assigned whenever a PSF is judged to support correct performance, but does not enhance performance (contrary to *good*) or make tasks easier to carry out than typically expected.
- *Poor*: A poor level of a PSF is detrimental towards the accomplishment of an objective (leading to the occurrence of a human error).

PSFs from this list will act as a basis for subsequent identification and characterization of safety critical contexts. Other factors will be interpreted as or mapped to PSFs from this list.

5.1.1.2 Identification of HFEs and safety critical context

Present work considers an HFE as a central starting point of analysis; (i.e. a top down approach) PSFs are linked to the said HFE in operational conditions. In a retrospective analysis from the accident/incident report, and for a prospective analysis from a task analysis-like approach, as discussed previously in [section 4.3.1.1](#). One of main reason behind this distinction is that the HFE identification is often made from a functional point of view in the classical sense of a *human error* (accomplishment of a safety related function). PSFs on the other hand are linked to operational conditions, and characterize the working environment in which the said human function is accomplished. EUAR HF and similar studies, which provide an operational safety-oriented analysis of human functions, can be used to find detailed analysis on possible PSFs. For a prospective analysis, such studies can be used to identify both HFEs and PSFs, whereas for a retrospective analysis the latter is more interesting.

As discussed previously, generic PSF lists contain anywhere from 8-15 PSFs [[Boring, 2010](#)], and represent a body of knowledge on factors to take into account when analyzing human reliability. However, not all PSF are present in a situation, or present in a degrading state at-least to merit a detailed quantitative analysis. This step is often done in most other models at the application step. However, since the quantitative model as described in the next subsection needs the identification of a (sub)set of PSFs. Thus, a characterization of the operational context is needed. More particularly for present works objectives, a context which can have a significant impact on human reliability.

A safety critical context is represented as a collection (set) of factors (PSFs) which impede a safe accomplishment of a human function (or an HFE). A presence of these factors is often linked to have considerable negative affect on human performance towards the accomplishment of said function. This section can either be made by the analyst, or a multiple source approach can be followed, the latter is detailed as in the context of present work. It follows the notion presented in [section 4.3.1.2](#), the objective there was to provide a relatively complete reference list of PSFs that an analyst can select by taking a union from all the sources. Here, characterization of a safety critical context is a refinement of the PSF list, to the operational condition. The straightforward usage is that an experienced analyst

Table 5.1 – PSF list with considered definitions and levels [Rangra et al., 2015a]

Performance Shaping Factor	Definition	Qualitative levels
Training	Have the correct knowledge to perform a job successfully and safely. Training might be needed to ensure skills are up to date and relevant, i.e., new procedures, different signaling systems, etc.	Good, Nominal, Poor
Experience	The accumulation of information and knowledge gained through interactions with the system and time spent in the work environment, this can be in same conditions (or same route). Aspects like, bad habits learned, etc. should also be considered in addition to the, positive aspects.	Good, Nominal, Poor
Communication	The ability of team members to pass information to each other and a shared understanding of the situation using, system status, read-outs, etc. e.g. misunderstanding, omission of, information, mistaken location, incorrect communication actions. Human-machine communication aspects are not included in this PSF.	Good, Nominal, Poor
Situational Awareness	The perception of the elements in the environment within a volume of time and space. The comprehension of their meaning and projection of their status in the near future. As a more general definition from [Endsley, 1995]: “ <i>Skilled behaviour, that encompasses the processes by which task-relevant information is extracted, integrated, assessed, and acted upon.</i> ”	Good, Nominal, Poor
Task Load (Workload)	The actual task demand assigned to a person in terms of the number and type of tasks (varying complexity, importance, etc.). Task load can also be impacted by unplanned or emergency events.	Good, Nominal, Poor
Time load (Workload)	Time required or allocated for one or multiple tasks; this time perception can affect worker stress beyond the stress of having too many tasks. Available time to complete a task particularly in the context of driving activities related to high speed trains (both detection and completion of the task).	Good, Nominal, Poor
HSI quality	An umbrella term to consider the quality of human system interface. The broad context here includes the procedures, appropriate information displayed to the human at appropriate time or in an adequate way. It includes most ‘Machine-based factors’ directly influencing human behaviour.	Good, Nominal, Poor

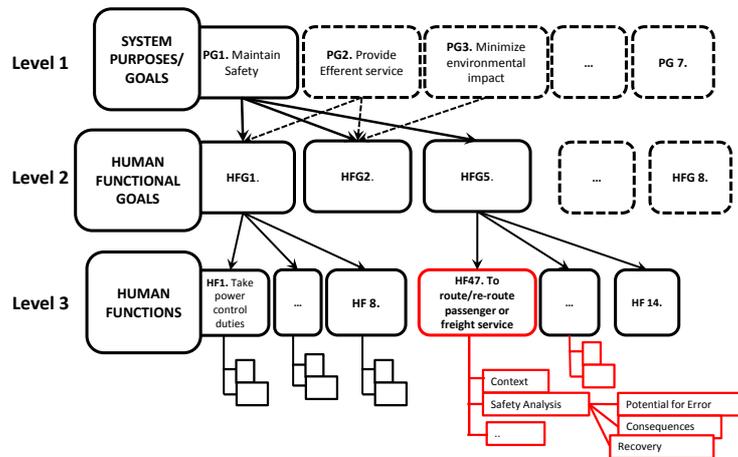


Figure 5.2 – The usage of EUAR HF study to identify the safety critical context, a refinement of the context related to an human function and in turn an HFE, is done using the data as marked in red.

uses the PSF list and selects the PSFs that need a quantitative analysis. If the analyst does not possess sufficient information concerning the operational condition, a multiple point of view approach as discussed in [section 4.3.1.2](#) can be used, in opposite sense, i.e. refinement of different sources of qualitative data by focusing on an HFE, and taking an intersection, that is selecting the common factors (PSFs) in the different sources.

The first source of data here is the human factors study EUAR HF. Since an HFE is identified, in order to use the HF study a human function which matches closely to the HFE under analysis is needed. Here, the HFE is the non-accomplishment (an error) of a human function. Safety relevant activities and analysis thereof provided in the study consist of discussion on respective conditions (PSFs) that could possibly lead to errors. To identify an error-causing context (as a set of PSFs) *potential for errors* and in some cases *potential for recovery* is used. Current approach only considers explicit statements (or the PSFs stated to be the ones with largest influence) to identify this context. Further, it can be considered that the factors with a negative connotation (*potential for error*, etc.), have a significant effect on the human while performing a said function. This, human function's error-causing factors are then interpreted in terms of PSFs from the PSF list (Table 5.1). This can be seen in the diagram shown in [Figure 5.2](#), for a human function the data marked in red is used. Error context (ECs) as discussed previously are also used. Since, ECs' application domain is nuclear; the definitions were matched to the PSFs in the PSF

list, to keep interpretation of PSFs coherent. The third source is the entire PSFs list, discussed in previous section as a standard base of factors. A refinement of the context linked to a given HFE is done by taking a union, i.e. common or recurring factors are selected. The preference in selecting the PSFs is as follows - EUAR HF study, PSF list and accident/incident report, or safety critical context/contextual factors identified by experts or analysts - are assigned higher importance than ECs. As evident, a higher preference is assigned to rail domain specific studies.

Thus, the proposition termed Human Failure Type Context – HFTC, is a qualitative construct which characterizes a specific HFE’s safety critical context. This safety critical context is represented as a set of PSFs specific to an HFE:

$HFTC_{HFE}$ where $HFTC_{HFE} = \{PSF_1, PSF_2, \dots, PSF_n\}$.

Example for illustration. An example will be used here (and in the following subsections) to illustrate the steps of PRELUDE. This approach is illustrated here by constructing the HFTC for a given HFE. As discussed before this HFE is assumed to be extracted from a scenario/task analysis process. Towards constructing the the HFTC a human function, the previously described multiple source approach is followed. First, the EUAR HF study is used. As shown in [Table 5.2](#) for a HFE, a reference to EUAR HF study (more specifically the spreadsheet accompanying it) is given as, *ERA/HFG i/j*, where *i* is the index of *human function goal* and *j* refers to the index of the specific *human function*. The definition of the human function *ERA/HFG5/47: Route/re-route passenger or freight service* matches closet to the definition of the HFE; the PSF list and the Error context are also used alongside. Common factors, interpreted as PSFs (underlined in [Table 5.2](#)) from the three sources are subsequently identified. This gives us $HFTC_{HFE} = \{\underline{Communication}, \underline{TaskLoad}, \underline{TimeLoad}\}$. This concludes the identification of the safety critical context of the HFE. All the variables of the quantitative model are thus identified. The second step towards quantification is detailed in the next section.

5.1.2 Quantitative part

Quantitative aspect of this proposition concerns with formalizing the evidence(s) to build human reliability model in VBS. That is, once an HFE is identified, the relations between the safety critical context (the PSFs) and the HFEs need to be defined. Presently, this evidence is obtained from expert elicitations. Information from multiple experts is elicited and combined, followed by a transformation to obtain the quantitative human reliability model. The following subsections describe each sub-steps in detail.

Table 5.2 – Defining Human Failure Type Context for the example HFE

HFE	Performance Shaping Factor identification	Human Failure Type Context (HFTC)
Wrong route/re-route of passenger or freight service	<p><i>Source 1.</i> ERA/HFG5/47: Potential for error – “Trains could be routed towards other traffic, incompatible infrastructure (59), engineering possessions, or close to engineering work at high speeds. Errors could be influenced by <u>time pressure</u> or <u>complexity</u> in track layouts in some situations and locations and problems with <u>communications</u> (It could be important to consider additional risks at shift changeover).”</p> <p><i>Source 2.</i> PSF list – Training, Experience, <u>Communication</u>, Situational awareness, <u>Task Load (Workload)</u>, <u>Time load (Workload)</u>, HSI quality.</p> <p><i>Source 3.</i> Error Contexts – EC1 = {Training, HSI quality, <u>Task load</u>, <u>Communication</u>, <u>Time load</u>}</p>	$HFTC_{HFE} = \{\text{Communication, Task Load, Time Load}\}$

5.1.2.1 The expert elicitation process

In the previous step the variables of the human reliability model, namely the HFE, and the safety critical context $PSF_i : PSF_i \in HFTC_{HFE}$ are identified and defined. This step aims to obtain data to build the qualitative relation between these variables. Configuration belief structures (or valuations) are used to formally define this relation. Towards this objective, a simple question-answer construct is used to capture focused domain knowledge from an expert. The HFE and the PSFs need to be contextualized for the expert. This is accomplished by using information from the EUAR HF study. For a retrospective analysis it can also include comments of investigators, chronology of events etc. However, statements which are inferences of the investigators on the factors and HFEs can potentially influence expert beliefs, and thus, should be avoided. Hence, this additional data hence aims to provide an expert a non-ambiguous description of the context.

This work’s application is concerned with the *true* state of an *HFE* and negative effects of a PSF, a *good* level is not considered in the questions. Furthermore, the experts are considered to have a complete knowledge concerning the questions asked. However, the model and transformation can account for a *good* level. This inclusion can be interesting to account for the effect of a PSF on suppressing the negative effect of another PSF. However, this requires some effort on expert elicitation, notably question structure, and combination. Nevertheless, it is not considered in the present quantitative proposition. Thus, only HFE being *true* or *false* ($\Omega_{HFE} = \{\text{true}, \text{false}\}$) and a PSF being *poor* or *nominal* ($\Omega_{PSF} = \{\text{poor}, \text{nominal}\}$) are used in the questions, and later transformations. A configuration belief structure can represent the conditional relations between multiple variables (PSFs and HFEs)

and their values.

The questions are formulated as conditional piece of information, that is the HFE's occurrence (*true*), based on the condition that a *poor* level of PSF is present. The text of the question, thus, forms a proposition, and the questionnaire aims to obtain a degree of confidence (as an expert's opinion) on the truthfulness of that proposition. The question is given below concerned PSF is formulated as:

Given the occurrence of a *poor* level of *PSF*, what do you think about the probability of the *HFE* being *true*?

The response is expected on a probability scale – number of times out of 10, 100, 1000, etc. the HFE is said to be *true*. That is, the probability that the human will fail to perform the safety critical task (i.e. the HFE is true), when the task is required to be performed, in the given conditions (PSFs). An expert can use descriptors – *d* or give a subjective probability value. The natural language descriptors or simply descriptors are taken from ATHEANA's elicitation process [Forester et al., 2004], where similar quantities are elicited. Nevertheless, current work uses them as they are given in [Forester et al., 2004], where these are defined as follows:

- 'Likely' to fail – 0.5 (5 out of 10 times the operator will fail to perform the given task)
- 'Infrequently' to fail – 0.1 (1 out of 10 times the operator will fail ...)
- 'Unlikely' to fail – 0.01 (1 out of 100 times the operator will fail ...)
- 'Extremely unlikely' to fail – 0.001 (1 out of 1000 times the operator will fail ...)

Thus, the response from the expert takes the form:

Given a *poor* level of PSF, the HFE is *true* with a probability of x .

Second set of questions is a PSF in it's nominal state and experts are questioned on the absence of the error. Essentially this represents the cases where the human is able to perform the task correctly, given that the PSFs are in a nominal state. This data can be relatively easy to obtain from other sources, since it is interested in the nominal state of PSFs, i.e. the situations where the PSFs are judged to not degrade human performance. Thus, empirical and or historical data can be used here. This question aims to complete the evidence, in terms of the values of the HFE (true and false), at least as far as considered in this work. This consideration is represented by the question:

Given the occurrence of a *nominal* level of all the *PSFs* what do you think about the probability of *HFE* being *false* ?

It is important that all the experts are given clear description of the variables and their definitions. The questions as formulated presently (a HFE-PSF pair) presents a generic and simpler context to visualize for the expert, making the elicitation process easier, both for the expert and the analyst. On the contrary, for example, multi-PSF questions might force experts to make conclusions on factors out of their domain of expertise. For example, a railway expert might be asked to comment on aspects of human cognition. This limitation is relevant because it is rather common in HRA assessment process to have multiple domain experts each with different expertise. Further, the combined effect of multiple PSFs, might lead to misunderstanding of the situation by the experts. Thus, more complex constructs need to be avoided when eliciting experts.

Example for illustration (cont.) For the HFE and its HFTC obtained after [Table 5.2](#), questions are given in [Table 5.3](#). The HFEs and PSFs are contextualized for the expert using data from EUAR HF study, more specifically in section *Personal and organizational goals, generic context and potential for error*. These choices are facultative and left to the discretion of the analyst. Present work aims at expert independence by eliminating direct expert interaction in the data collection process. It was ensured that there was no interaction among the experts during the expert data collection process, and experts do not have access to each other's responses. And therefore their responses to a question are treated as independent pieces of evidence.

5.1.2.2 Combination of expert data

This section follows the introductory discussion on combination rules presented in [section 2.2.2](#). While mathematically aggregating the data, the hypothesis of the data aggregation method needs to be respected, and the choice remains with the analyst. This choice should be based on the experts and data to be combined.

Thus, when constructing the quantitative human reliability model, PRELUDE offers to the analyst, a choice of the combination method to use. The choice depends mainly on the hypothesis attached to said rules which can be applied to evidence at-hand. This paper also provides an illustration of what different rules can be used, their hypothesis, and the results. Thus, all combination rules are used in the case study to illustrate some aspects of expert data combination, notably conflict. Some comments are also presented in the case study.

Present work's objective is to understand and demonstrate, when and what method to use, based on the underlying hypothesis. All of the five combination

Table 5.3 – The HFE’s description and relevant context description and question statements

HFE and questions	Context description and question statements
Wrong route/re-route of passenger or freight service	<p>A signaller was <i>Not able to (Route / re-route passenger or freight service).</i>;</p> <p><i>Personal and organizational goals</i> – “To respond to scenarios that require trains to be re-routed or travel to a different (unplanned) destination.”;</p> <p><i>Generic context</i> – “Ensure train services can continue operations during engineering / maintenance work, enable engineering trains to get to the work area, provide adequate routing plans, this routing of trains could be planned prior to the work; require short term (re-)planning where there is limited notice of engineering work. . .”;</p> <p><i>Potential for error</i> – “Trains could be routed towards other traffic... Errors could be influenced by time pressure or complexity in track layouts in some situations and locations, and problems with communications (It could be important to consider additional risks at shift changeover).”</p>
Question 1.	Given the occurrence of a <i>poor</i> level of <i>Task Load</i> , what do you think about <i>HFE</i> being <i>true</i> ?
Question 2.	Given the occurrence of a <i>poor</i> level of <i>Communication</i> , what do you think about <i>HFE</i> being <i>true</i> ?
Question 3.	Given the occurrence of a <i>poor</i> level of <i>Time Load</i> , what do you think about <i>HFE</i> being <i>true</i> ?
Question 4.	Given the occurrence of a <i>nominal</i> level of <i>all the PSFs</i> what do you think about <i>HFE</i> being <i>false</i> ?

methods are thus used in this paper (ref. [section 2.2.2](#)). A summary of their hypothesis and assumptions are given in [Table 5.4](#).

Table 5.4 – Combination rules and their hypothesis

Combination methods	Hypothesis and it's manifestation
Arithmetic average	All experts are equally reliable. The data received is thus, given equal weight.
Weighted average	A differentiation between experts' domain knowledge is made. This manifests as weights assigned to the evidence received from each expert.
Independent consensus or majority vote	There is a single correct answer to the question. Therefore, the answer which has the highest frequency (relative) is chosen. However, if no clear majority amongst the values/descriptors is found, an arithmetic average is used.
Dempster's combination rule	All the experts are equally reliable and evidences are independent. It is associative, commutative but non-idempotent. It essentially weakens the disagreement and strengthens the agreements in terms of conflict in the elicited values.
Yager's rule	This rule assumes that all the experts are reliable and the evidences are independent. It is quasi-associative, commutative but not idempotent. Contrary to previous case the conflict manifests itself as uncertainty.

Thus, for a proposition say X (what the question aims to measure) it is considered that the expert is fully sure of the response, as " X is exactly x and only x "; where X can have as values $\{x, \bar{x}\}$. Thus, an expert's belief is represented by a *bpa*. The value of this *bpa*, say b , is a quantitative expert belief (a subjective probability) on the said proposition. Each expert's response is then modeled as a complementary belief structure. This goes to state that, for each expert, belief about the value of X being x is b and exactly b . Therefore, the belief of $X = \{\bar{x}\}$ is $1 - b$. This is then modeled as two focal sets with the associated *bpa* values. The belief structure in [Equation 5.1](#) gives the considered representation of expert data.

$$\begin{aligned}
 m(\{x\}) &= b \\
 m(\{\bar{x}\}) &= 1 - b \\
 m(\Omega_X) &= m(\{x, \bar{x}\}) = 0
 \end{aligned}
 \tag{5.1}$$

Finally, after combining the data, a single response (a quantitative value) for each question (PSF-HFE pair) is obtained; this is used in the next section to complete

the VBS model construction.

5.1.2.3 Transformation

Combination of data gives a single piece of evidence per PSF-HFE pair. This evidence is a combined quantitative measure of the experts' opinion on each question's proposition. These measures and propositions are used for constructing the configuration belief structure for the VBS model. To construct this belief structure appropriate transformations of the data are needed. First, each question's proposition represents a conditional piece of evidence, viz. a conditional belief of a state of an HFE given a state of a PSF. This conditional belief must be transformed appropriately to accommodate it in the dynamic part of VBS. Secondly, the simpler questions asked to the experts need to be combined in this step to obtain the complete human reliability model.

For the first part, the relations between variables should be represent as valuations or joint belief. Thus, Smets's rule [Xu and Smets, 1996] is employed. It propose to transform a conditional piece of evidence into a joint belief structure (or a **de-conditioning**). It represents a conditional relation between two variables A ($\Omega_A = \{a, \bar{a}\}$) and B ($\Omega_B = \{b, \bar{b}\}$), such that the belief about B is known only when the actual value of A is known [Xu and Smets, 1996]. This transformation is defined as follows defined as given in Equation 5.2.

Given the conditional evidence if $A = a$ then $B = b$ with a $bpa = x$.

The rule is represented by a belief structure defined on $E : \Omega_E = \Omega_A \times \Omega_B$,

such that: the focal set $\{(a, b)(\bar{a}, b)(\bar{a}, \bar{b})\}$ is assigned a $bpa = x$,

and the focal set Ω_E is assigned a $bpa = 1 - x$ (5.2)

After using the rule given in Equation 5.2, for every question, an initial belief structure is obtained which relates a particular PFS and a HFE with a bpa value. This initial belief structure is defined on the frame $\Omega_\Phi = \Omega_{PSFi} \times \Omega_{HFE}$.

Second, $HFE = \{\{true\}\}$ and $PSF = \{\{poor\}\}$ is the minimal explicit information in a question's proposition. On the other hand, the VBS model quantifies using a set of valuations. For present work this valuation (the configuration belief structure) relates HFE and all the $PSF \in HFTC_{HFE}$. That is, it reasons with the HFE and with all the PSFs in an given safety critical context (i.e. $HFTC_{HFE}$). Effectively it is defined on the frame $\Phi : \Omega_\Phi = \{\Omega_{HFE} \times \Omega_{PSFi} | \forall PSFi \in HFTC_{HFE}\}$, i.e. all the states of all the PSFs in an HFTC and the relevant HFE. Hence a transformation is

needed to obtain the complete final configuration belief structure. Thus, a **vacuous extension** (section 2.2.2, Equation 2.4) is performed on the initial belief structure (if needed), to generate the intermediate belief structures for each question. This transformation, thus entails a *vacuous extension* of a question's initial belief structure giving the intermediate belief structure.

Finally, all the questions' propositions are represented as their respective intermediate belief structures. These intermediate belief structures for each question defined on the same frame. In order to obtain a complete VBS model, which represents quantitatively the relation between the HFE and it's safety critical context (HFTC) a **final combination** of the questions is needed. The independence constraints are respected while eliciting the experts, as discussed previously, hence, they can be combined using Dempster's rule. Further, as can be seen in Table 5.4, if there are n PSFs in an HFE's HFTC, there are $n + 1$ number of questions. Thus, this final combination is given in Equation 5.3, for all the questions to obtain the final configuration belief structure of the complete VBS model.

$$m = mQ1^{\Omega_{HFE} \times \Omega_{PSF_1} \times \Omega_{PSF_2} \dots \times \Omega_{PSF_n}} \oplus mQ2^{\Omega_{HFE} \times \Omega_{PSF_1} \times \Omega_{PSF_2} \dots \times \Omega_{PSF_n}} \oplus \dots \oplus mQn + 1^{\Omega_{HFE} \times \Omega_{PSF_1} \times \Omega_{PSF_2} \dots \times \Omega_{PSF_n}} \quad (5.3)$$

This gives us the final configuration belief structure concluding the construction of the VBS model, from the the simple questions (section 5.1.2.1). More generally, multiple PSFs and an HFE (PSF-PSF-...-HFE) questions are not asked from the experts in present work, they can very well be implemented in the current proposed approach (by adequately changing the vacuous extension). Nevertheless, in our configuration belief structure there is always a component of an HFE. Pure PSF-PSF are currently not considered. However, mathematically speaking, as is the case with the work discussed before [Groth and Mosleh, 2012b], a PSFs influence on another PSF can be modeled using an intermediate belief structure, which, for example can be a belief structure between two or more PSFs, which then links to other PSFs and HFEs. Nevertheless they are not considered in present work. The next section describes the transformation for the example HFE introduced previously.

Example for illustration (cont.) This transformation is explained using the HFE's questions given in Table 5.3. Given that multiple experts are elicited, the data is combined using a combination rule (ref. section 5.1.2.2) the results should be obtained as a probability value associated with each question. This transformation step works onwards from that combined data, it is assumed (for illustration

purposes of this example) that the final probability values are obtained as follows:

- Combined probabilistic response for Question 1. : 0.05
- Combined probabilistic response for Question 2. : 0.2
- Combined probabilistic response for Question 3. : 0.001
- Combined probabilistic response for Question 4. : 0.95

Question 1 from Table 5.3 can be written as: if $Task\ Load = \{poor\}$ then $HFE = \{true\}$, with the combined probabilistic response representing being 0.05. Further, as discussed in expert elicitation, it is considered that $\Omega_{PSF} = \{poor, nominal\}$ and $\Omega_{HFE} = \{true, false\}$, the following abbreviations are used to refer to the PSFs: TaL for Task Load, TiL for Time Load, C for Communication.

This proposition is **de-conditioned** using the rule in Equation 5.2 giving two initial belief structures (bpas):

$$\begin{aligned}
 & mQ1^{\Omega_{HFE} \times \Omega_{TaL}}(\{(poor, true)(\overline{poor}, true)(\overline{poor}, \overline{true})\}) \\
 &= mQ1^{\Omega_{HFE} \times \Omega_{TaL}}(\{(poor, true)(nominal, true)(nominal, false)\}) = 0.05, \text{ and} \\
 & mQ1^{\Omega_{HFE} \times \Omega_{TaL}}(\Omega_{HFE} \times \Omega_{TaL}) \\
 &= mQ1^{\Omega_{HFE} \times \Omega_{TaL}}(\{(poor, true)(nominal, true)(poor, false)(nominal, false)\}) \\
 &= 0.95
 \end{aligned} \tag{5.4}$$

Here, since $\Omega_{PSF} = \{poor, nominal\}$ it can be considered that for a PSF $(\overline{poor}) = (nominal)$. Similarly initial belief structures can be obtained for Questions 2 and Question 3. Each of these initial belief structures contains two elements per set. Note that, Questions 4 however, represents a relation between all the PSFs and the HFE. It is therefore interpreted as: if $(TaskLoad, Communication, TimeLoad) = \{nominal, nominal, nominal\}$ then $HFE = \{false\}$. The initial belief structure thus in this case contains four elements per set. Thus, all of the focal sets using equation 5.4 and combined probabilistic responses as the *bpa*'s are given below. These initial belief structures (that is the respective focal sets and bpas) for the questions obtained after de-conditioning using equation [Xu and Smets, 1996], are given below:

Question 1. represented as $mQ1^{\Omega_{HFE} \times \Omega_{TaL}}$, which gives first focal set as: $\{(poor, true)(nominal, true)(nominal, false)\}$, with a *bpa* = 0.05; and second focal set as: $\{\Omega_{HFE} \times \Omega_{TaL}\}$
 $= \{(poor, true)(nominal, true)(poor, false)(nominal, false)\}$ with a *bpa* = 0.95.

Question 2. represented as $mQ2^{\Omega_{HFE} \times \Omega_C}$ which gives first focal set as: $\{(poor, true)(nominal, true)(nominal, false)\}$ with a $bpa = 0.2$; and second focal set as:

$\{\Omega_{HFE} \times \Omega_C\} = \{(poor, true)(nominal, true)(poor, false)(nominal, false)\}$ with a $bpa = 0.8$.

Question 3. represented as $mQ3^{\Omega_{HFE} \times \Omega_{TiL}}$ which gives first focal set as: $\{(poor, true)(nominal, true)(nominal, false)\}$ with a $bpa = 0.001$; and second focal set as:

$\{\Omega_{HFE} \times \Omega_{TiL}\} = \{(poor, true)(nominal, true)(poor, false)(nominal, false)\}$ with a $bpa = 0.999$.

Question 4. represented as $mQ4^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}}$ which gives first focal set as: $\{(nominal, nominal, nominal, false), (poor, poor, poor, false), (poor, poor, poor, true)\}$ with a $bpa = 0.95$ and second focal set as:

$\{(nominal, nominal, nominal, false), (poor, poor, poor, false)$
 $(nominal, nominal, nominal, false), (poor, poor, poor, true)\}$ with a $bpa = 0.05$.

As can be seen in the first column of above equations questions 1, 2 and 3 are defined on $\Omega_{HFE} \times \Omega_{TaL}$, $\Omega_{HFE} \times \Omega_C$ and $\Omega_{HFE} \times \Omega_{TiL}$ respectively. Thus, these initial belief structures need a **vacuous extension**. However, Question 4 is already defined on the frame $\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}$. It thus, does not need a vacuous extension. Thus, a vacuous extension is performed for questions 1, 2 and 3. This process is detailed below:

For Question 1's proposition the obtained bpa $mQ1^{\Omega_{HFE} \times \Omega_{TaL}}$ is defined on the frame $\{\Omega_{TaL} \times \Omega_{HFE}\}$, whereas in the present VBS model, the configuration belief structure is defined on the frame $\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}$. Thus, a vacuous extension is performed by performing a cross product of the elements of the focal sets obtained after obtaining the initial belief structures, and frame of *Task load* and *Communication*, this extension is given as follows:

$$\begin{aligned}
 & mQ1^{\Omega_{HFE} \times \Omega_{TaL}}(\{(poor, true)(nominal, true)(nominal, false)\}) \\
 & \quad = mQ1^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}}(A), \text{ where} \\
 & A = \{(poor, true)(nominal, true)(nominal, false)\} \times \{\Omega_C \times \Omega_{TiL}\} \\
 & \quad = \{(poor, true)(nominal, true)(nominal, false)\} \\
 & \quad \quad \times \{(poor, nominal)\} \times \{(poor, nominal)\} \\
 & \quad = \{(poor, true, poor, poor)(poor, true, poor, nominal) \\
 & \quad \quad (poor, true, nominal, poor) \dots\} \quad (5.5)
 \end{aligned}$$

Here, A is the focal set for the *bpa* $mQ1^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}}$. The complete focal set thus obtained is defined on the frame $\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}$. It thus, contains four elements per set, instead of previous two. Similarly, 5.5 is done for all of the questions' focal sets, with appropriate frames to obtain intermediate belief structures for each of them. Finally all the questions are **combined** using Dempster's rule. For the example this is given in Equation 5.6.

$$m1 = mQ1^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}} \oplus mQ2^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}} \oplus mQ3^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}} \oplus mQ4^{\Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}} \quad (5.6)$$

This gives us the final configuration belief structure for the example HFE.

5.1.3 Quantification and sensitivity analysis

In this final step of PRELUDE, the VBS model takes data on the PSF(s) to quantify the HFE (the variable of interest). This data is represented in the direct belief structure(s), called here **the direct evidence**. It can come from a given operational or an accident context, and is assigned by an analyst. It simplifies the usage and eliminates the aspects of subjectivity on the choice of PSFs and their affects (given the VBS model is constructed beforehand) during the analysis process. As a natural second step, a feedback for the analysis undertaken can be performed. The objective of this feedback in PRELUDE is to aid decision making by allowing an analyst to perform a diagnosis on the individual PSFs. This is interpreted as suggestions for improvements to be made in terms of the PSFs to reduce the probability of the occurrence of the HFE. This is what is called here **a sensitivity analysis**. Both of these two steps are described in this section.

5.1.3.1 Assigning the direct evidence and quantification

Since, current application deals with a retrospective analysis, this evidence is obtained from relevant accident analysis statistics. For present work [Kyriakidis, 2013] is chosen for its relevance to the domain and availability of data matching the current need. That is the number of times a PSF was one of the causal factor(s) given there was a human involvement in an event (accident, severe accident, etc.). Furthermore, if a PSF is identified as a cause of an accident, it can be safely assumed to be in a *poor* state, or a state which degrades human performance in general. Also, they arrive at R-PSF and their occurrence frequencies after merging the PSFs (and subsequently their occurrence frequencies) in multiple steps (definitions,

categorization, and threshold). That is, we can consider that after these multiple levels of combinations, the resulting PSFs and their frequencies are sufficiently independent, to be considered as such in the model.

Thus, in present case a direct belief structure represents proposition that a PSF level is *poor*. The quantitative measure on this evidence is obtained as: number of times a PSF is reported to be a direct cause vs. total number of events. Thus, the proposition is represented by the focal set $\{(poor)\}$ and the quantitative measure on this evidence assigns the bpa: da_{PSF} (for direct assignment to a PSF). If PSF was not in a *poor* state, it is everything except *poor*. That means, it is *nominal*, since for current work $\Omega_{PSFi} = \{(nominal, poor)\}$ such approximations are sufficiently conservative for present work. This gives us the proposition and evidence thereof as $m^{PSF}(\{(nominal)\}) = 1 - da_{PSF}$. Similarly for all the PSFs, and thus, direct belief structures for the VBS model are obtained. These direct evidences come from independent and reliable sources. That is, all the evidence obtained from accidents are independent (1.5 PSFs identified per accident report) and reliable (accident/incident investigation reports) [Kyriakidis, 2013]. Thus, in this case Dempster's rule can be employed.

After defining all the direct belief structures, BFM (Belief Functions Machine) [Giang and Shenoy, 2003] is the software used to combine the direct and configuration belief structures and marginalize for HFE. The results obtained by marginalization, i.e. projection on Ω_{HFE} are given in Table 5.6. The obtained results are represented as upper and lower probability bounds (as described in section 2.2.1, Equation 2.3). Since, human reliability analysis and also present work is concerned with an HFE being *true* these quantification results are represented in the form of an interval, given as:

$[Pr_{inf}(HFE(true)), Pr_{sup}(HFE(true))]$, as described in section 2.2.1.

5.1.3.2 Sensitivity analysis

To perform the sensitivity analysis the problem is set-up, by modifying and then using the VBS model as follows:

1. Modification of the VBS: An HFE is assigned a direct belief structure where the focal set of $\{(true)\}$ is a bpa value equal to 1, i.e. the error has occurred.
2. Marginalize for each PSF: mainly the *poor* state in the $HFTC_{HFE}$.

The obtained marginal for each PSF is combined with the direct evidence thereof (again using Dempster's rule). It is essentially an updating of evidences. It includes the combination of a prior (obtained from experts – the configuration belief

structure and the HFE being *true*), and a posterior (assigned for the application, represented by direct belief structure) evidence. Mathematically the marginal is obtained for each power set of the values of the PSF, similar to what was obtained for the HFE quantification in [Table 5.6](#). However, towards safety objectives only the state of PSF under analysis is kept (i.e. *poor*), other focal sets and their bpas are not discussed. The results of the sensitivity analysis are presented as a bar graph. The marginal obtained for the *poor* state of a PSF are presented in percentage form. In other words each bar represents the relative percentage contribution of a PSF, as being *poor* with a certain *bpa*, with the given relational (HFE-PSF) and situational evidences. This is interpreted as the contribution of a PSF towards causing an ‘error’. It is to be noted, that this choice of states of the variables (*poor* and *true*) is driven by the objectives of the current analysis. This operation can be performed on any variable and any of its state(s). The percentage values are used to ranks PSFs in terms of their contribution towards causing (the PSF *poor* leading to HFE *true*) the HFE. This makes it possible to establish a priority ranking, towards improvements in PSFs needed for effective gains in operational safety and to identify PSFs on the other end of this list, which can be ignored.

Example for illustration (cont.) For the HFE from [Table 5.2](#), the domain of interest is $\Phi = \{HFE, TaL, C, TiL\}$. Where, *HFE* is the HFE under analysis, and the PSFs – Task Load (*TaL*), Communication (*C*) and Time Load (*TiL*). Their respective frames are defined as $\Omega_{HFE} = \{true, false\}$, and for each of the PSFs as $\Omega_{PSF} = \{nominal, poor\}$. The relation between the PSFs and HFEs, is defined by the configuration belief structure obtained after [Equation 5.6](#), represented graphically as *m1* in [Figure 5.3](#).

Here, *m1* is defined on the frame $\Omega_{\Phi} = \Omega_{HFE} \times \Omega_{TaL} \times \Omega_C \times \Omega_{TiL}$. The other bpas *m2*, *m3* and *m4* contain evidence on the variables *TaL*, *C* and *TiL* respectively. As discussed before, they are direct belief structures and are used to represent data on the PSFs. In [Figure 5.3](#) shows the graphical model with the direct and configuration belief structures titled what they contain. The direct assignment for Task Load is represented by the diamond shaped node – *DataOnTaL*. It contains two focal sets $\{nominal\}$ and $\{poor\}$, and respective *bpa* values as $m^{\Omega_{TaL}}\{(nominal)\}$ and $m^{\Omega_{TaL}}\{(poor)\}$. The actual values are direct evidences obtained from accident statistics, as discussed in the previous section. Similarly for Time Load and Communication, these direct evidences are given in [Table 5.5](#)¹.

For the example HFE being *true*, the results in [Table 5.6](#) are represented in the

¹Task Load and Time load are defined as a single R-PSF in [[Kyriakidis, 2013](#)]; whereas current PSF list they are different, thus, its frequency is divided equally amongst the two.

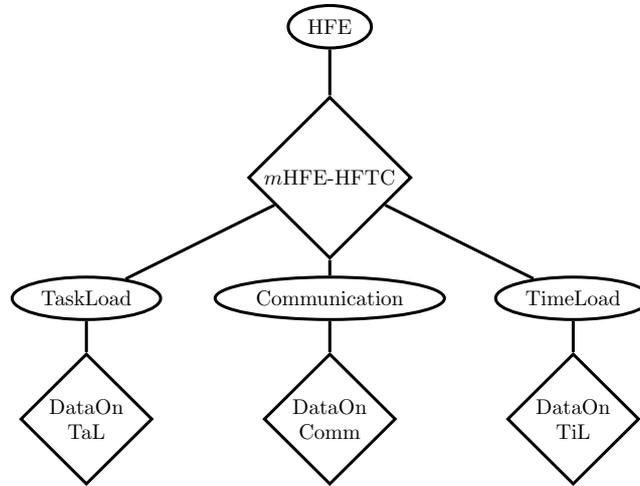


Figure 5.3 – Graphical representation of the example HFE’s implementation in VBS.

Table 5.5 – Direct belief structures for HFE: from the R-PSF equivalent as identified focal sets and associated *bpa* values.

Cause from [Kyriakidis, 2013]	R-PSF Identification frequency Total number of accidents – da_{PSF}	PSF and representation in Figure 5.3 vs. m_2, m_3, m_4	$m^{PSFi}(\{poor\}) = da_{PSF}$	$m^{PSFi}(\{nominal\}) = 1 - da_{PSF}$
Workload, Time pressure, Stress.	58/1676 = 0.0173	Task Load (m_2)	0.0346	0.965
Communication, Teamwork.	228/1676=0.136	Communication (m_3)	0.136	0.864
Workload, Time pressure, Stress.	58/1676=0.0173	Time Load (m_4)	0.0346	0.965

Table 5.6 – Marginalization results for example HFE on Ω_{HFE} .

Values of the example HFE	<i>bpa</i> on the HFE’s values obtained after marginalization
$m^{\Omega_{HFE}}\{true\}$	0.00005
$m^{\Omega_{HFE}}\{false\}$	0.94981
$m^{\Omega_{HFE}}\{true, false\}$	0.05014

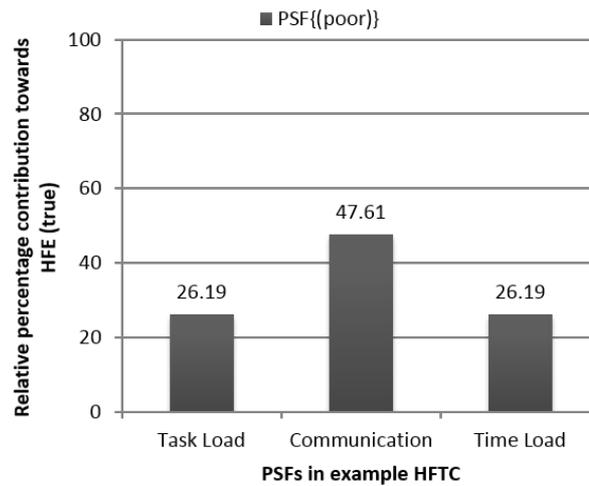


Figure 5.4 – Sensitivity analysis results for example HFE and associated PSFs

form of an interval, given as:

$[Pr_{inf}(HFE(true)), Pr_{sup}(HFE(true))] = [0.00005, 0.05019]$. The sensitivity analysis results obtained for the example HFE are shown in **Figure 5.4**. It can be seen that Communication has clearly a higher contribution than the other two PSFs. Thus it can be concluded that improving the aspects of Communication should be the priority. This interpretation will be discussed in details in the case study.

It can be remarked here that for questions of the example HFE, expert valuations (*bpa* values) are different (i.e. combined probabilistic response for a question, as given in **section 5.1.2.3**). On the other hand, the direct belief structures in **Table 5.5** are the same for the case of Time Load and Task Load. Now, as seen in **Figure 5.4**'s, the feedback seems to reflect direct evidence, as it states Task Load and Time Load (for their poor level) to be equally likely contributors towards the HFE. However, it does not seem to reflect the differing expert valuations. This is due to the fact that the configurations belief structure's focal sets are symmetric, as obtained by the transformation approach from **equation 5.6**. This leads to the marginal obtained in the first step of sensitivity analysis being same across all the PSFs. This is a constraint stemming from the simpler questions, and by extension the transformation approach asked to the experts, as a question with more than one PSFs will generate a non-symmetric focal set in the configuration belief structure. But then again an expert might find it difficult to respond to such questions, however, other data sources (such as simulator experimentation) can be used. Nevertheless, this concludes the quantitative proposition employing the VBS model.

5.2 Case study

This section presents the application of PRELUDE on a recent catastrophic high-speed railway accident's scenario. In this accident human error was concluded to be as one of the primary causes. Data (factors, events, etc.) are taken from the official investigation report [[Comisión de investigación de accidentes ferroviarios, 2014](#)]. This work neither aims to nor can achieve the detailed and exhaustive qualitative analysis provided in the official investigation report. Here, the prime motive is to demonstrate the usage and application of PRELUDE as a retrospective analyses.

The usage of the PRELUDE in this case study is demonstrated by employing the three steps as shown in [Figure 5.1](#). This application process is generic, the way in which each step is conducted depends upon the purpose of the analysis. Step 1 follows the traditional sense of defining the scope of the analysis, and analyzing an accident scenario to identify the HFEs and related PSFs, to characterize a safety critical context. For a prospective approach this can be a procedure and operational context. Step 2 puts PRELUDE's quantitative propositions from [section 5.1.1](#) and [section 5.1.2](#) into action – elicitation of data from experts, and combination and transformation thereof. Finally, in Step 3, the quantification data for application (direct evidence) is input, and the results of quantification and sensitivity analysis are presented and commented on.

5.2.1 Step 1. Qualitative part: HFE and PSF(s) identification

As a pre-cursor to the application of PRELUDE, this step defines scope of the analysis to limit the problem-set. Main considerations include: type (retrospective), and detail (procedures, and human actions or functions). Since current scope is limited to analyzing the accident scenario and demonstrating key aspects of the proposition, thus, only the immediate HFEs and PSFs which are direct causes of the accident are analyzed.

The report provides a detailed and chronological account of noteworthy events which led to the accident in question. It is thus, used to identify the HFEs. A reverse task-analysis approach is implemented, where the starting point is the immediate safety critical events involving a human (HFEs) in the accident report. Further, as needed detailed operating procedures and the signaling principles were consulted from the national regulation documents, such as directive guidelines which contain procedures requiring 'passive and immediate obedience' from a human actor. These directives can be considered to have a higher priority than for example, non-regulatory or non-normative guidelines such as "good practices in

driving". The sources and how they are used is given as follows:

- Accident investigation report – identification of HFE (non-accomplishment of a task or function) and description.
- National regulations – detailed description of the procedures, previously identified task is part of.
- Human factors studies (EUAR HF) – identification of safety critical situation mapping of human function and previously identified HFE.

Finally, for the identified HFE, excerpts from the report and the relevant procedures given in Table 5.7. The plot in Figure 5.5 aims to show the HFEs (annotated in yellow-red ovals) in chronological occurrence of the accident scenario. The horizontal axis represents distance from the accident point in meters and vertical axis represents speed in km/h.

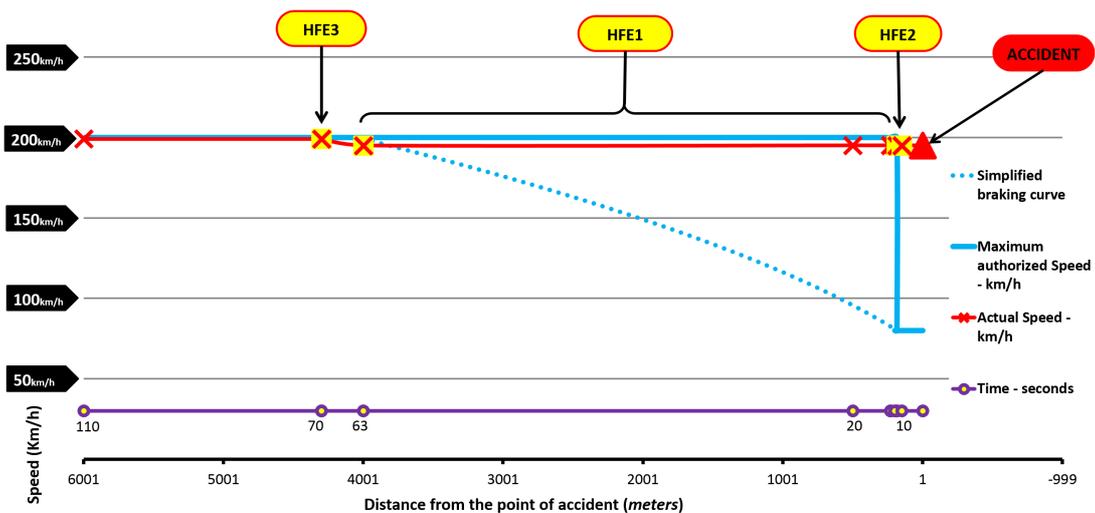


Figure 5.5 – Accident scenario: speed of the train vs. distance from the point of accident. The data points (cross marks) are other events as identified in the investigation report and the HFEs. Also a time scale is given to represent the time duration of the analyzed scenario

Further, the safety critical context of the respective HFEs needs to be identified. This activity follows steps of section 5.1.1 identify the PSFs and subsequently $HFTC_{HFE}$ for every HFE. The EUAR HF study is here used for additional identification of the PSFs, using a mapping of HFEs to *human function*. A *human function* which matches closest to the HFE under analysis is identified from the EUAR HF study. EUAR/HFG4/35 defined as "Maintain appropriate speed" which entails the *personal and organizational goal* as "To ensure movements at a speed that is safe for the vehicle in the current conditions and in accordance with the timetable."

Table 5.7 – HFE and relevant procedures from the accident investigation report and national regulations

Identified Human Failure Event from the accident report	Source and procedure/task description
HFE1. Ineffective communication	<i>Accident investigation report</i> – "... a contributing cause was an absence of attention of the train driver ... after answering a phone call..." No strict procedures/rules (except a guide to good practice in driving.)
HFE2. Not respecting the speed signals (in schedule book/table of speeds)	<i>Accident investigation report</i> – "... not respecting the prescribed maximum speed allowable by infrastructure, as established in the tables of maximum speeds mentioned in the train Schedule book..."
HFE3. Not reducing speed in time	<i>Accident investigation report</i> – "driver should identify the reference (point) to initiate the braking and to reduce the speed." <i>Regulations</i> – "The driver shall endeavor to recognize the signs (signals) as far as possible and do not lose interest in their observation as (long as) it (train) has not crossed them."

This definition matches with both the HFE2 (Not respecting the speed signals in schedule book or table of speeds) and *HFE3* (Not reducing speed in time) as identified in Table 5.7. In addition to the sources used in section 5.1.1.2 (i.e. EUAR HF study, PSF list, ECs) extracts from accident investigation report are also used to account for accident scenario specific PSFs. It gives a retrospective account of casual factors that influenced the human towards those errors.

For example, it states – "the train driver (the human actor) did not brake (the activity required of him) because of distraction (the factor)." Hence, the context is dealing with the train driver performing the human function/task of braking; the act of not braking (non-accomplishment) is the HFE. Here, "distracted" becomes a factor of interest for a safety critical human function that is the act of braking, that is it's $HFTC_{HFE}$.

It is important to note that, the use of accident report augments the HFTC's construction and is application specific (qualitative part, dashed document box Figure 5.1). All the identified PSFs and their sources, and the recurring PSFs (underlined based on the preference presented in section 5.1.1.2), and finally the HFTCs are given in Table 5.8.

From a functional perspective, the act of braking and respecting a speed signal forms one single function. In the EUAR HF study the scope of *human function* (HFG4/35) entails both HFE2 and HFE3. Present work considers them separately

Table 5.8 – Identification of PSFs for defining HFTC for HFEs from the accident scenario

HFE	Performance Shaping Factor identification	Human Failure Type Context ($HFTC_{HFE}$)
HFE1	<p>EUAR/HFG5/53 – <u>Communication</u>; PSF list – Experience, <u>Communication</u>, Situational awareness, Task Load (Workload), Time load (Workload), HSI quality; Accident investigation report – <u>Communication</u>; Error Context= {Training, HSI quality, Task load, <u>Communication</u>, Time load}</p>	$HFTC_{HFE1} = \{\text{Communication}\}$
HFE2	<p>EUAR/HFG4/35 – Training (skill), <u>Experience</u> (local knowledge); PSF list – Training, <u>Experience</u>, <u>Communication</u>, <u>Situational awareness</u>, Task Load (Workload), Time load (Workload), <u>HSI quality</u>; Accident investigation report – <u>Situational Awareness</u> ('lack of attention'), <u>HSI quality</u> ('lack of regulation on track-side fixed preventive signaling'); Error Context – {<u>Situational awareness</u>, <u>Experience</u>}, ; {Task load, <u>HSI quality</u>, Time load, <u>Situational Awareness</u>}</p>	$HFTC_{HFE2} = \{\text{Experience, Situational Awareness, HSI quality}\}$
HFE3	<p>EUAR/HFG4/35 – Training (skill), <u>Experience</u> (local knowledge); PSF list – Training, <u>Experience</u>, <u>Communication</u>, <u>Situational awareness</u>, Task Load (Workload), Time load(Workload), HSI quality; Accident investigation report – <u>Situational awareness</u> (as dissonant 'cognitive location'), <u>Time Load</u> (refer Figure 5.5); Error Contexts – EC2 = {<u>Situational awareness</u>, <u>Experience</u>} ; {Task load, HSI quality, <u>Time load</u>, <u>Situational Awareness</u>}</p>	$HFTC_{HFE3} = \{\text{Situational Awareness, Time Load, Experience}\}$

and they were quantified as separate HFEs. An appropriate granularity level in accordance with analysis objectives and thus, appropriate source of data must be selected. Current approach, takes such data from multiple sources to precisely identify the PSFs involved. Such, details might not be otherwise observed, if a single point of view is taken.

5.2.2 Step 2. Quantitative part: Expert elicitation, data combination and transformation

Three experts with different domain expertise were consulted. Their combined expertise covers human factors engineering, railway signaling, BFT, and safety and reliability aspects of the railway domain in general. Such a variety of domain knowledge is in-line with what is advised by other such expert-data based methods. Further, independent elicitations were carried out. The experts were sent the questions and related context detail. In [Table 5.9](#) shows the questions and descriptions for HFE2. Similarly structure was followed for other HFEs identified in Step 1.

Some experts chose to respond using the descriptors, whereas some felt comfortable with giving directly probability values. The data thus obtained, from the three experts (A, B, and C) are given in [Figure 5.6](#). The data thus obtained, for each question from the three experts A, B, and C are given in [Figure 5.6](#). Subsequently, data from each expert for each question was combined using different combination rules ([Table 5.4](#)). For weighted average (WA) combination, for demonstration purposes a choice was made to give a higher weight to expert with experience in the railway industry (expert C). Thus, the following normalized weighting factors were chosen: 0.2 for expert A and B, and 0.6 for expert C. The combined values for each question thus obtained is also given in [Figure 5.6](#).

Separate belief structures were generated ([section 5.1.2.3](#)) for each of the five combination methods. That is for each combination method used, different *bpa* was generated for each HFE's VBS model. Expanding on the discussion of [section 5.1.2.2](#), here we briefly comment on the different combination rules to demonstrate the difference. The figure [Figure 5.6](#) shows in the form of the grouped bar plots for each question – the expert data (first three bars) and the data obtained after different combination rules (latter bars in the same group). The expert data for *question 1.1* is commented here, similar comments can be made for other questions. It can be observed that expert B and C give the same probability value, whereas expert A gives a significantly lower probability value. This discussion is from the perspective of what the experts say (the probability values they give) and the combined data that is obtained. Following remarks can be made:

Table 5.9 – Context description and questions for HFE2 sent to the experts

HFE and questions	Context description and question statements
<i>HFE2</i> . Not respecting the Speed Signals (in schedule book).	Definition of HFE from Table 5.7 , and more details. "... speed change from 220 km/h to 80 km/h ... track-side information is a marker indicating a change in maximum permitted speed." <i>General remarks</i> – The signaling system/ATP in place does not protect against over speeding in the case of permanent maximum speed changes. The train driver is wholly responsible for this action.
Q1. <i>Experience</i> - <i>HFE2</i>	Given the occurrence of a <i>poor</i> level of <i>Experience</i> , what do you think about <i>HFE2</i> being <i>true</i> ?
Q2. <i>Situational Awareness</i> - <i>HFE2</i>	Given the occurrence of a <i>poor</i> level of <i>Situational Awareness</i> , what do you think about <i>HFE2</i> being <i>true</i> ?
Q3. <i>HSI quality</i> - <i>HFE2</i>	Given the occurrence of a <i>poor</i> of <i>HSI quality</i> , what do you think about <i>HFE2</i> being <i>true</i> ?
Q4. all PSF - <i>HFE2</i>	Given the occurrence of a <i>nominal</i> level of all the PSFs what do you think about <i>HFE</i> being <i>false</i> ?
Additional information	PSFs and their definitions in $HFTC_{HFE2}$ (definitions of <i>Situational Awareness</i> , <i>HSI quality</i> , <i>Experience</i> from Table 4.4). Answering aid/instructions: The response is expected on a probability scale, i.e. how many times out of 10, 100, 1000, etc. do you expect an HFE to be <i>true</i> , that is the operator failing to do the required task. Natural language descriptors can also be used, they are defined as follows: <i>Likely</i> , 0.5 (5 out of 10 times the operator will fail); <i>Infrequently</i> , 0.1 (1 out of 10 times); <i>Unlikely</i> , 0.01 (1 out of 100 times); <i>Extremely unlikely</i> , 0.001 (1 out of 1000 would fail).

- A weighted average is different than simple average since it enriches an average by taking into account the expertise of an expert (higher weight).
- A vote tends to account for essentially what majority of experts say, irrespective of their expertise (weight) or the difference between the values themselves (high or low conflict).
- In Dempster's rule the difference between the values is interpreted as conflict (k). Subsequently, Dempster's rule manages this conflict by normalizing it. The higher the conflict, the more normalization is performed. There is a conflict in this case, thus, a value smaller than previous combination rules is obtained. Furthermore, non-idempotent nature of Dempster's and Yager's rule (section 5.1.2.2) gives some interesting results for question 1.2., 2.4. and 3.4. That is even though all experts give the same probability values (independent consensus) the combined value is different.
- Yager's rule computes the conflict similarly, but treats it as an uncertainty instead of normalizing it. Therefore, a value higher than Dempster's rule is obtained. It may be noted that the version of Yager's rule (dynamic) used here is quasi-associative (section 5.1.2.2). That is the order in which the expert data is combined has an influence on the combination result. Thus we see that, for question 3.1. vs. 3.2. Yager gives different results (1.99E-03 vs. 1.09E-03).

Here the frames for $HFE2$ and $HFE3$ were considered same as the example HFE in item 5.1.3.2, that is $\Omega_{HFE} = \{true, false\}$, and for PSFs as $\Omega_{PSFi} = \{nominal, poor\}$. Configuration belief structures were defined and transformed the same as other questions (following steps of section 5.1.2.3). For all the HFEs the modeling in VBS is thus complete. The implementation in VBS is given for HFE1 in Figure 5.7, HFE2 in Figure 5.8 and HFE3 in Figure 5.9. It can be noted that the VBS models thus constructed are not specific to the case study. It can be used wherever similar HFEs and PSFs are identified. The quantification results and feedback/sensitivity analysis results are presented and discussed in the next step.

5.2.3 Step 3. Quantification data and results

Finally, in the last column of Table 5.10 the data for quantification, i.e. direct evidence, for respective direct belief structures are obtained from accident statistics data (explained in section 5.1.3.1). This data for quantification is assigned to all the PSFs of the case study (data on PSFs - as seen as the in Figure 5.7, 5.8 and 5.9.

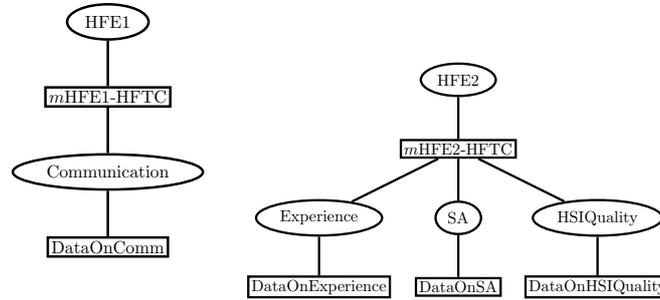


Figure 5.7 – VBS model of HFE1 **Figure 5.8 – VBS model of HFE2 and its HFTC.**

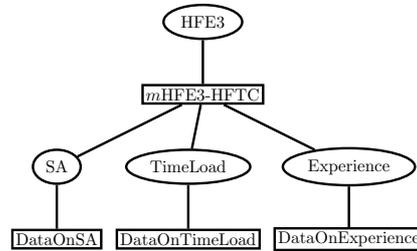


Figure 5.9 – VBS model of HFE3 and its HFTC.

Table 5.10 – Direct belief structures (focal set and bpa) for the PSFs in the case study, identified similar to as in Table 5.5 from the R-PSF equivalent

PSFs used in the case study	bpa for focal set $m^{PSFi}(\{(poor)\})$	bpa for focal set $m^{PSFi}(\{(nominal)\})$
Communication	0.136	0.864
Experience	0.0787	0.9213
Situational Awareness	0.144	0.856
HSI Quality	0.1885	0.8115
Time Load	0.0173	0.9827

Table 5.11 – Top: Middle of the probability interval for the variable $HFE1$ value of interest ($true$) as obtained after combining data using different combination rules. Bottom: The pairwise distance metric d_J from equation 2.7, between the bpa s obtained for $HFE1$.

Middle of interval	mAverage	mWeighted Average	mVote	mDempster	mYager
$HFE1(true)$	0.113	0.117	0.124	0.071	0.108

bpa obtained from combination rule	mAverage	mWeighted Average	mVote	mDempsters
mWeighted Average	0.797	-	-	-
mVote	0.797	0.799	-	-
mDempsters	0.817	0.819	0.821	-
mYagers	0.797	0.799	0.801	0.82178

Each of the combination rules used gives a different result. However, one combination rule can be chosen to obtain the necessary results i.e. an interval for each state of the variable of interest (an HFE). Nevertheless, towards PRELUDE’s guide on the choice of combination methods (in addition to discussion in section 5.1.2.2) a brief discussion is presented here. This comparison is different than what was discussed in the previous section, where the combination of expert data and immediate results were discussed. Further, in this case the combination rules are compared at the level of end-results, that is for the HFE. Only $HFE1$ are discussed here, similar metrics can be computed for other HFEs.

The Table 5.11 gives the middle of the interval for the variable of interest $HFE1(true)$ as described in section 2.2.3. In this case the $true$ value of HFE is chosen.

It can be remarked that the middle of the interval for all the combination rules are similar except for Dempster’s rule. This is because of the way in which it manages conflict is different than Yager’s rule, and others which do not manage conflict explicitly (see section 2.2.2).

The distance metric d_J can take a maximum value of 1 [Jousselme et al., 2001]. Thus we can observe in Table 5.11 that all of the bpa s have high distance values between them, that is they are dissimilar pairwise. However, relatively speaking each bpa is at equal (maximum distances between two bpa s 0.821 and minimum 0.797) distances from each other.

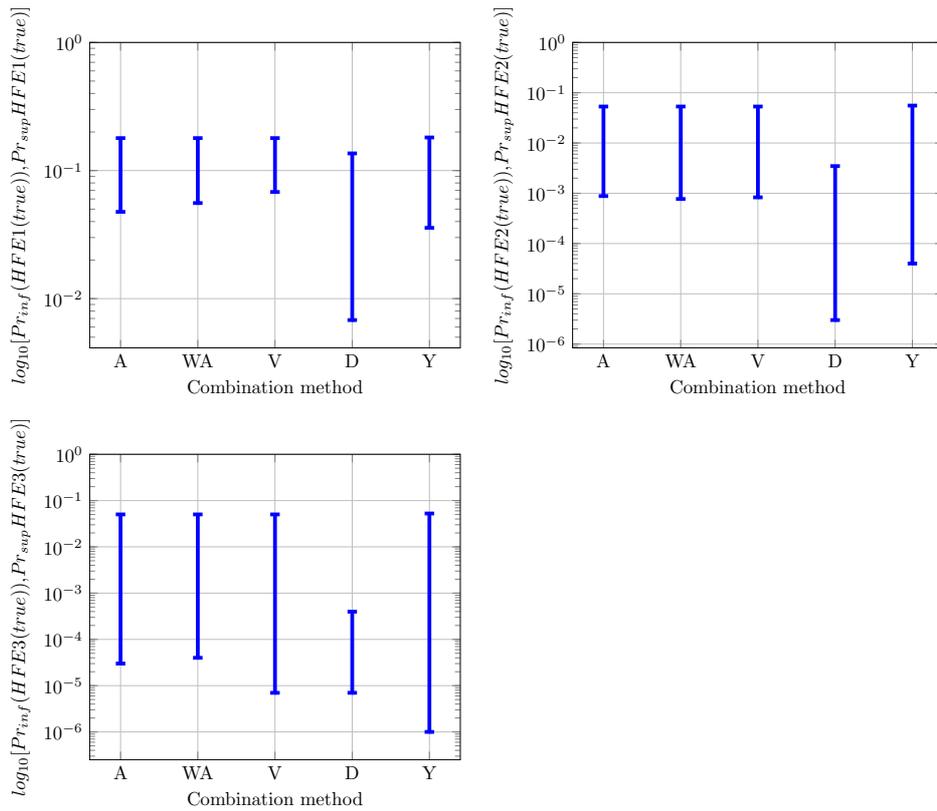


Figure 5.10 – Lower and upper bounds for an HFE’s true state, different models are built for when combining expert elicitation using average (A), weighted average (WA), vote/independent consensus (V), Dempster’s (D) and Yager’s (Y) combination rules

In this case, the distances are not significantly informative, they do not provide sufficient information to single out particular rules or a pair thereof which is dissimilar/similar to the others. Nevertheless, such a metric can be useful when it is difficult to make a choice of combination rule only based on the underlying hypothesis.

The quantification results are obtained in the form of upper and lower probability bound. The value of variable of interest is an HFE’s true state, that is the focal set $\{(true)\}$. The quantification results are given in [Figure 5.10](#). The x axis marks the combination method used and the y axis is the probability value in \log_{10} with an axis maximum of 0.2. The line-plot in y axis presents the log of lower and upper probability bound for the variable HFE value of interest $true$. The lower bounds for first three combination rules was of the order of 10^{-3} , for latter two i.e. BFT-based, however it was 10^{-6} . This is generally the case with human error probabilities in other HRA methods.

The results of $HFE1(true)$ have higher upper bounds due to the higher values obtained (an average of 0.35) obtained from the experts. For both $HFE2(true)$ and

$HFE3(true)$ in terms of interval size and lower bounds, a clear distinction must be made between BFT-based (Dempster's and Yager's), and other combination rules. This is due to the latter with implicit and the former with explicit handling of conflict. Weighted average implicitly removes conflict by weighing values obtained from one expert (expert C, see [Figure 5.6](#)) more than others. Nevertheless, the intervals are not completely disconnected since conflict is relatively less; if it was higher significant changes could have been observed. Therefore, simpler combination methods are as effective when conflict is low or ignorable. In the case where ($HFE2$ and $HFE1$) lower bound in Dempster's rule is smaller than Yager's; this is due to the presence of high conflict in elicited data (ref. [Table 5.6](#)). In Yager's hypothesis, the conflict amongst expert's values is put in uncertainty, instead of normalizing the lower bound. Thus, in both the cases of $HFE2(true)$ and $HFE3(true)$, Yager's rule gives a larger interval size as compared to Dempster's results.

Thus, even though these combination rules provide an accurate and a formal representation of uncertainty, in some cases (a high conflict) it might make decision-making difficult.

Towards keeping this choice open, PRELUDE methodology remains adaptable and only needs conditional quantitative data on a PSF-HFE pair. The transformation ([section 5.1.2.3](#)) manages the rest to construct the VBS model. This also ensures that explicit considerations of epistemic uncertainty (in the human error relational model) are uniform irrespective of the combination rule used. Subsequently, for $HFE2$ and $HFE3$ a sensitivity analysis (following the steps in [section 5.1.3](#)) is undertaken. $HFE1$ is not analyzed further because it contains only one PSF.

It can be noted that for each expert data combination method, different VBS models (configuration belief structures) are generated, for space constrains it is not possible to discuss all of them here. Hence, only average combination rule is selected for both HFE and only the *poor* level of a PSF is considered. The [Figure 5.11](#) shows the obtained sensitivity analysis results. They are presented in the form of a relative percentage value, interpreted as a relative contribution towards causing an HFE to be *true*. It can be seen that, *HSI quality* for $HFE2$ and *Situational Awareness* for $HFE3$, have the highest relative contribution.

Hence, the sensitivity analysis indicates that the occurrence of these HFE could be reduced by improving aspects of *HSI quality* and *Situational Awareness*, as first priority.

The results effectively are also indicative of the reality (the accident investigation report) and expectations of experts. $HFE2$ can be classified as a checking error [[Embrey, 1986](#)]; for a checking error the quality or source of information (i.e. HSI

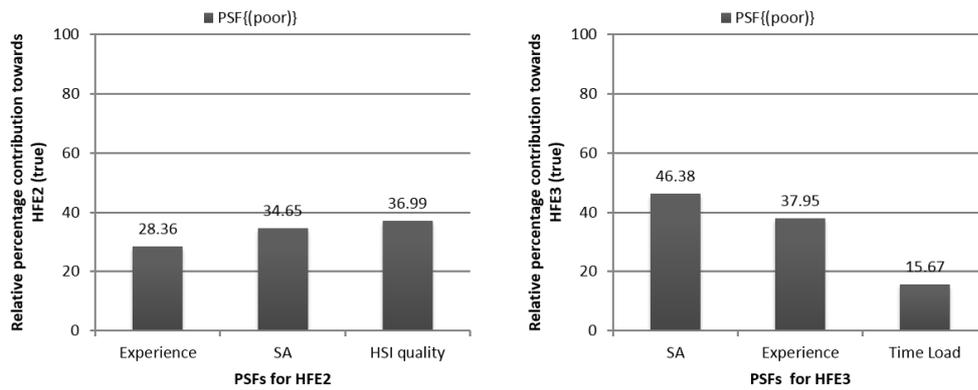


Figure 5.11 – Sensitivity analysis results of PSFs for the context of *HFE2* and *HFE3*

quality) can be judged to be relatively critical than other factors. This corresponds to the results seen in [Figure 5.11](#).

Thus, it can be remarked that the model gives results corresponding to conclusions from the human error mechanisms.

These results are also reaffirmed in the accident analysis report. The investigators concluded that the availability of right information at the right time was indeed one of the issues. Subsequently, a change of the placement of track-side speed markers was recommended by the accident investigation report [[Comisión de investigación de accidentes ferroviarios, 2014](#)]. In particular, special considerations (more track-side markers, gradual speed changes, etc.) are made for places where a high change in ceiling speed (200 km/h to 80 km/h, blue line in [Figure 5.5](#)) is required. For *HFE3*, PRELUDE's analysis concludes that *Situational Awareness* has the highest relative contribution. This also, corresponds to the concluding statement of the report that absence of attention was one of the main causes of the accident.

Thus, the proposed quantitative model gives results coherent with expert and domain knowledge.

5.3 Conclusions

PRELUDE methodology in its current state is developed for and applied to a case study for the railway domain. It may very well (with moderate efforts) be applied to other domains. A completely generic HRA model is rare and such models are often inspired by the notions of previous domain-specific HRA models.

It entails an expert system built using concepts from BFT in VBS which formally defines the casual elements of a human reliability model. Decision-making

capabilities are provided by using probability intervals and sensitivity analysis to establish a priority ranking amongst the PSFs for a given safety critical context. PRELUDE is centered on Performance Shaping Factors specific for railway needs. PSFs and human factor studies are employed for an adequate representation of human factors and operational safety concerns. It was also demonstrated as how to employ human factor study towards HRA objectives. The expert system's graphical representation as a VBS allows for an easy representation of variables and their relationships, and, thus simplifies the usage by analysts and non-experts in the mathematical framework.

A formal combination and transformation proposal is used to build the elements of quantitative human reliability model from expert data. This approach to formally model human failure events as a function of the PSFs in evidential networks (VBS) offers a novel perspective for human reliability quantification. The expert data can be replaced by an empirical source, with the condition that it is in the form of a conditional probability, at the very least HFE-PSF (the questions currently asked of the experts) or HFE-multiple PSFs. Conditional in terms of the variable HFE and PSF, the states thereof poor or nominal, true or false, are not a limitation to BFT framework. Thus, if the empirical data can be formulated as a (conditional) belief structure, it can be integrated into the VBS model currently proposed.

Although, when combining evidences (direct and configuration belief structures) the resulting frame is the product of the respective variables' frames, there the size of the product space may create a computational bottleneck. That is to say, more granular variables with larger frames (i.e. multiple states of PSFs and HFEs) are difficult to combine, compared to e.g. multiple variables (multiple PSFs and HFEs) with less number of states (smaller frames). Some rules for combination of expert elicited data are also contrasted. A relatively straightforward middle of the interval comparison, and a standardized distance metric are used. These metrics, as used in the context of present work allow comparing the results obtained by using different combination rules in a quantitative manner. It allows the experts to identify the appropriate choice of combination rule to use or justify the choice.

If particularly conflicting expert opinions are considered, it is observed that some methods result in larger intervals than others. Thus, the choice thereof is left to the analyst; nevertheless, the usage of such methods needs further investigation. Sensitivity analysis results were used to establish a priority rank towards improvements in PSFs needed for effective gains in human reliability. Although it is more interesting if multiple PSFs and multiple HFEs are modeled in the same EN, this might lead to non-evident sensitivity analysis results. The proposition's implementation on a retrospective analysis of a real-world railway

accident demonstrated the usage of the methodology. Once the VBSs models are built the implementation is relatively straightforward, and interval provides accurate representation of uncertainty in data (if any) and easy decision making. The results obtained correspond to theoretical and expert expectations. Thus, VBS offers an adequate framework in its utility towards newer generation of human reliability methods.

System level, risk-based inferences, and taking contribution of positive aspects of PSFs and human actions need to be appended to PRELUDE towards a robust methodology. This is aimed at further simplifying the decision-making capabilities, and performing a holistic analysis of a human error. Feedback or remarks of the experts on the structure of questionnaire can also be obtained to improve the elicitation process, mainly to ensure that the analyst and experts have the same understanding towards assuring the accuracy of obtained data. Validation using simulator data and sensitivity analysis can be employed towards immediate verification objectives. Further, usage of empirical data from operational simulators to reinforce expert knowledge and to validate the methodology needs to be explored further.

Feasibility study of PRELUDE with data from simulator experimentation

Contents

6.1 Introduction	139
6.2 Experimental protocol using an operational simulator to obtain human reliability data	140
6.2.1 Simulator set-up: description of the ERTMS operational Simulator	141
6.2.2 The simulation environment	143
6.2.2.1 Rail track and procedures	144
6.2.2.2 Description of scenario runs	144
6.2.2.3 Explanation and basic training	148
6.2.3 Output data sources and analysis	149
6.2.3.1 Objective data: source and calculation of scores	149
6.2.3.2 Subjective data: source	154
6.2.4 Experimental campaign and preliminary discussion of the collected data	155
6.2.4.1 Objective data	157
6.2.4.2 Subjective data	158
6.3 Using data obtained from experimental protocol in PRELUDE	163
6.3.1 Pre-analysis: A classification of subjects	163
6.3.2 Objective data usage: combination with expert data – input to PRELUDE Quantitative part	166
6.3.3 Subjective data usage: Retrospective identification of PSF self estimation – input to PRELUDE qualitative part	172

6.4 Discussion	176
6.5 Conclusion	178

6.1 Introduction

In the previous chapter, PRELUDE used quantitative data from experts, and qualitative data from some human factors and accident analysis studies. Both sources can be augmented using empirical data. This chapter presents a feasibility study of the PRELUDE methodology with data from simulator experimentation. This chapter firstly proposes a protocol to obtain human reliability data from a simulator. This protocol is aimed at gathering both *subjective* and *objective* data, with final objective to support the PRELUDE methodology. We start by detailing the experimental protocol, the simulator set-up, and the data sources. An experimental campaign which was carried out is also presented, with a brief discussion of the data that was collected.

The second part of this chapter presents the method to input this data to the PRELUDE methodology. The objective data is treated as conditional data on human performance given the state of PSFs. In particular, we aim to use the objective data to combine with expert data and input to the VBS model. Subjective data is used to identify PSFs in a retrospective (empirical human performance data to safety analysis) approach. Note that the subjective data referred this chapter is not the same as subjective probability elicited from the experts in the previous chapter; it is a self-assessment by the experimental subjects.

Similar to PRELUDE's application the current work focuses on railway signaling. In particular the cab-driving aspect of ERTMS at its center, thus an ERTMS operational simulator is used. The simulator set-up used in the present work is similar to training simulators often used in the railway industry to instruct and train railway operators. Furthermore, the use of standard objective criteria and subjective questionnaires provide an easily repeatable, and adaptable to another domain of application.

Figure 6.1, shows the overview of the PRELUDE methodology, with the particular parts that this chapter aims to support marked in dashed borders. That is first, the quantitative formal relations between the PSFs and HFEs as modeled in the VBS. And secondly the data for the identification of the safety critical context of an HFE, that is the PSF(s) that are implicated in context of an HFE.

To clarify the difference between different actors we give the following definitions for this chapter. *Experts* refers to the domain experts in-general, similar to the experts that were consulted in PRELUDE's case study. *Operators* is when we refer to professionals in a work setting, e.g. a train driver. *Subjects* refers to the persons who take part in the experimental campaign, they may or may not be operators. Operators can be used as experts if they are involved in the expert elicitation

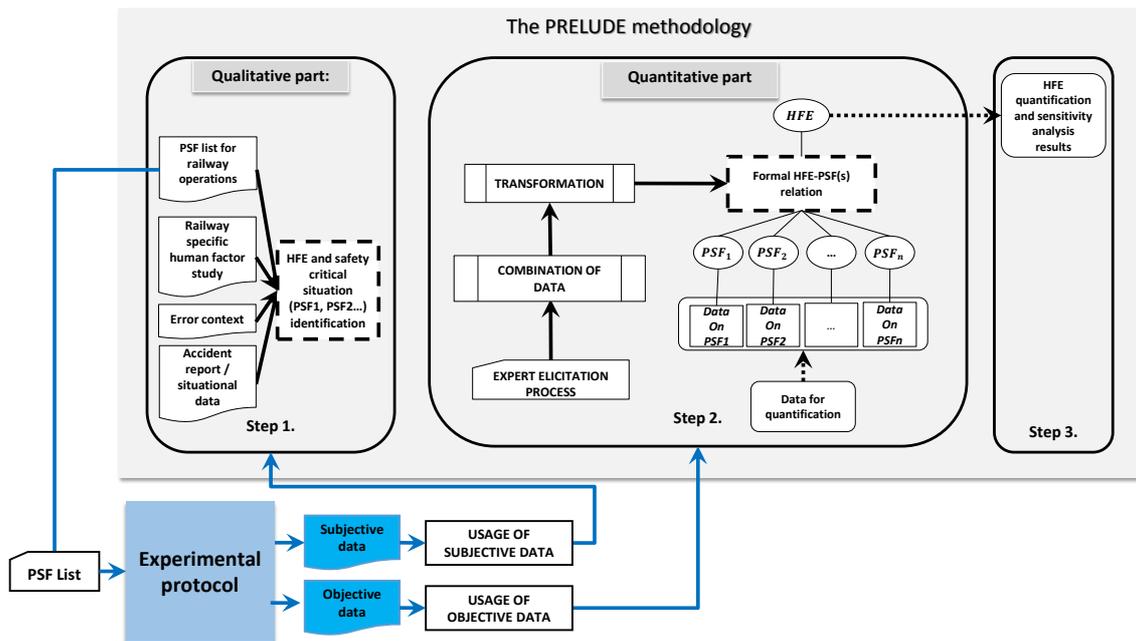


Figure 6.1 – Overview of PRELUDE’s methodology with the data requirements in dashed boxes, and experimental protocol’s inputs and outputs

process. If they are used in the experimental campaign they are addressed as subjects. That is the data that is obtained from an actor defines their role. Also expert data is not treated with the subjective data (more details are given while discussing subjective data analysis).

6.2 Experimental protocol using an operational simulator to obtain human reliability data

In-line with HRA objectives present work is interested in conditions detrimental towards human performance essential to ensure safe operation. In particular this work aims to capture the effect of PSFs and their states on human performance, towards modeling that effect in a human reliability model. To note that unlike most other approaches which observe human failure rates (frequency of errors) in multiple scenarios, this work is interested in how different PSF’s states affect that frequency. This protocol describes how and what data to obtain from the simulator towards supporting the PRELUDE methodology.

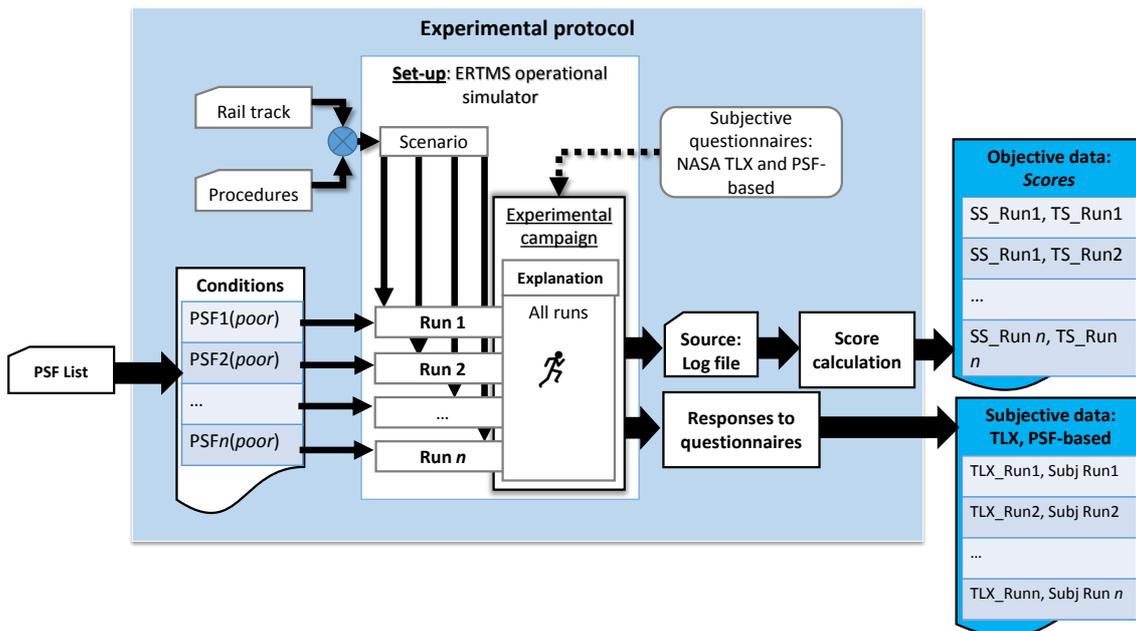


Figure 6.2 – Overview of the experimental protocol showing the inputs: the PSF list from PRELUDE, and the outputs: the objective and subjective data

As seen in top left of **Figure 6.2**, we start with a track and set of procedures for a scenario created on the operational simulator. The next subsections will detail the contents of this protocol: starting with the simulator set-up detail of the hardware used, the simulation environment: the scenario, the runs etc., and the source of the objective and subjective data. And finally we briefly present the experimental campaign that was carried out, and preliminary results.

6.2.1 Simulator set-up: description of the ERTMS operational Simulator

The present work has ERTMS at its center. An explanation of the levels and technical details were presented in **section 3.4.1**. In the case study of PRELUDE we analyzed a railway accident. This chapter also remains in the context of railway, but focuses in particular to the ERTMS signaling context. Thus, Heudiasyc laboratory’s ERTMS operational simulator ¹, hereafter referred to as the simulator is used. Similar set-ups are used to instruct and qualify railway operators. It allows creation of scenarios and running real-time simulations of trains running on tracks under ERTMS/ETCS supervision. It consists of various modules, namely: traffic management module (creating, managing and launching scenarios), driving module (EuroCab Simulator, the ETCS Driver Machine Interface - DMI) and 3D module (view of the track

¹<https://ferroviaire.hds.utc.fr/>

from a train driver's cabin). This simulator is compliant with specifications version (baseline) 2.3 of ERTMS supplied by the EUAR.

The simulator used in this work allows creation of scenarios and running real-time simulations demonstrating how trains can be run on tracks under ERTMS/ETCS supervision. It consists of different modules managing different components of the simulation (ETCS Track Editor, train dynamics module, EVC (European Vital Computer), ERTMS/ETCS DMI module, interlocking simulation, route map, RBC simulation, etc.). We can use these functionalities separately or all at the same time depending upon the scenario and the kind of data that we aim to observe. It contains multiple interconnected PCs for creating and simulating real train operational scenarios, these machines and their functionalities are described below:

- One *TRAFFICSIMU*, shown with a orange rectangle in [Figure 6.3](#) is a Linux server-grade desktop is used to control the whole simulator rig. Working under the Linux environment, it is connected to the OPSIMUx (see below) modules via a Local Area Network. This forms the core of the simulator functionality and also performs other support systems such as database management, creation of infrastructure, creation of scenarios, launching scenarios, scripts, etc. When running simulation scenarios this machine also functions as a *control center*, displaying RBC state, interlocking display, assigning routes and other dynamic parameters of the simulator.
- Four laptop PCs called *OPSIMUx* shown in yellow pentagon in [Figure 6.3](#) are each running a Linux environment to implement the human-machine interface (more specifically ETCS DMI - Driver Machine Interface) and the train control interface. As shown in [Figure 6.3](#) the ERTMS level diagrams) of a train equipped with ETCS (DMI, EVC, etc.).
- Two desktop machines - *UTC3Dx* shown in red rectangles in [Figure 6.3](#) with a Windows environment that allows a view of the trains in their natural environment (the terrain), track-side signalling, and other track characteristics, during a simulation. They can each be associated to a OPSIMU PC, thus for a given simulator session the combined: 3D and DMI view can be obtained for a

The main functionality of this simulator is to provide an environment, both physical and technical for train operation simulations, in the presence of the ERTMS regulations. It allows using the ETCS on-board interface. The ERTMS on-board functionality is used in the present work to provide a train driver's interface to carry out the simulations. As stated before, for a given scenario a maximum of 2



Figure 6.3 – ERTMS/ETCS Operational Simulator set-up: a picture, and its architecture listing the functions and different machines

complete set-ups, that is a DMI and corresponding external environment view can be used. Each of these two posts include a ETCS DMI and an 3D display of the track on a desktop machine display. These machines can be seen in [Figure 6.3](#) the *OPSIMU1* with *UTC-3D1* and *OPSIMU2* with *UTC-3D2*. Both of these configurations can be run together or separately, by modifying the parameters of the simulation. Various parameters can be changed such as: initial parameters of ERTMS/ETCS implementation (RBC, levels, etc.), there were some modifications made in the set-up of scenarios to allow for an additional variation in human performance.

Mainly a delay in application of automatic brakes (Service and Emergency Brake) in case of over speed. This allowed for an overspeeding the subjects can over speed more than what would normally be possible in a strict ETCS operation.

6.2.2 The simulation environment

The simulation environment is a track section: which is a mix of ERTMS supervision (Level 2, 1 and 0) to have a short but varied operational context. Some example tasks/procedures are: observe fixed track-side signals, respect indications on the DMI, respect timetable, etc. A *scenario* is thus defined as a train driver driving on a given track, performing associated tasks/procedures.

Since we are concerned with degraded conditions, a scenario is then modified to account for such conditions that lead to a higher probability of human error. This we interpret as PSFs' state *poor*, for example distraction from main task, bad communication, etc. As a reminder, a PSF is *nominal* if it is judged to support correct performance, and it is *poor* when is detrimental to performance needed towards the accomplishment of an objective [Table 5.1](#). Thus, we aim to simulate (for the subject) such degraded conditions. We were inspired by real world cases, standard practices and PSFs definitions. To keep data relevant to HRA, the scenario

runs take into account PSFs critical towards safe operation [Rangra et al., 2015b]. This is the PSF list that was also used in PRELUDE's quantitative part. Since there are multiple PSFs to consider therefore multiple *scenario runs* were defined. Each scenario run aims to simulate a different PSFs in a degraded state. For each scenario run, raw objective data is saved from the simulator, i.e. section 6.2.3.1. Subjective data as detailed in section 6.2.3.2 is also saved. This objective and subjective data pair is saved for each run by each subject. This data collection and analysis criteria is described in the following sections.

6.2.2.1 Rail track and procedures

As said before the track's signaling is a mix of ERTMS supervision (Level 2, 1 and 0, in that order). The ceiling speed profile generated by the ETCS on-board system is given in Figure 6.4, along with some procedures and signal boards that the subjects encounter during the simulation. These speeds are shown to a driver and require an absolute respect thereof. This track section and speed profile remain the same for all of the runs.

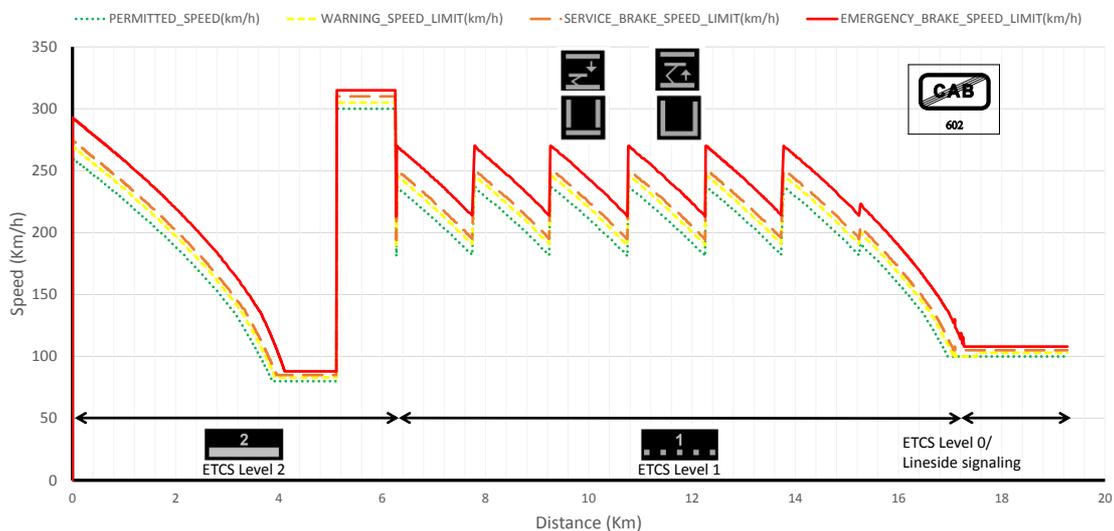


Figure 6.4 – ETCS on-board generated speed profile of the track used in present work

6.2.2.2 Description of scenario runs

This section details the background and creation of the *scenario runs*. To keep the data interpretation, and implementation of the protocol simple the following considerations are made:

- A given scenario is modified corresponding to only one PSF's state (bottom left [Figure 6.2](#)).
- This modification aims to create the effect (for the subjects undertaking the scenario) of a specific context, that is a degraded state of a PSF.
- The degraded state of the PSF is limited to a specific instantiation of the generic definition of the PSF for the real scenario.

Thus, in each case a PSF is selected and the scenario is appropriately modified. This gives the different scenario runs, we use the terminology in the remainder of this paper: Run 1, Run 2, etc. as also seen in [Figure 6.2](#). These runs are explained as given below, mainly for each run how a scenario was modified, what PSF it aims to make in a degraded condition.

Run 1. Training and Experience: This is the first run, it aims to simulate conditions of a poor *experience and training*. This being the first run, it is considered that the subject has a poor experience. Since this is the first run it is considered that the subjects have a low experience with driving a train, or more particularly the scenario that we have created. We consider that the level of experience equates to the time spent on the simulator. Furthermore, a state of poor training is simulated by not explaining (not included in the pre-simulation explanation and training) two crucial procedures required in the scenario. To note that after completing this run, the subjects receive an explanation of these procedures. These particular procedures are described as follows:

- ETCS Level 1 to Level 0 transition: a change in driving mode [section 3.4.1.1](#).
 - A yellow blinking icon appears on the DMI, the system demands an acknowledgment from the driver in the form of clicking/pressing the blinking icon.
 - If no acknowledgment is received in a predefined time interval (generally 5 seconds), the train stops (emergency brake application).
- Loss of power zone:
 - The driver needs to cut off the power by setting traction to zero by sliding a control of the train and re-accelerate. If not done, the train loses power and coasts till the said action is performed.
 - This can also lead to a time delay, although not as much since there is no active braking involved when coasting.

Run 2. Communication: The communication considered in this scenario is the communication between a train driver and a controller (human-human). A poor communication condition is simulated by interrupting the subjects with multiple messages, and to communicate the required information within stipulated time. Thus, during this run multiple messages in text form are displayed on the DMI. These messages either give the driver some information or ask the driver to relay some information to other human actor (the administrator). They appear based on train position (i.e. same for all of the scenario runs). All of these locations are fixed (against train position). Further, the subject needs to do the corresponding communication task, within a given period of time (fixed time period). Different human actors are to be addressed corresponding to a message received. They are described as follows:

- CONTROL : as the signaller at the control center
- AGENT : as the official in the train (a conductor or guard)

The different types of messages that a subject can receive are listed below:

- Perform a readout of the information on the DMI - *current speed, speed limit, distance to LRBG, current time.*
- *Information from CONTROL:* information messages from the control center, e.g. "the train will arrive at terminal 1".
- *Communicate to CONTROL:* the driver is asked to communicate some information to the control center, e.g. current speed and speed limit, state of next signal, etc.
- *Communicate to AGENT:* the driver is asked to relay additional information, e.g. if the train will arrive at the required time or not, if not, how much time delay (vs. the given timetable), etc.
- *Communicate to AGENT:* other auxiliary information such as the arrival terminal of the train.

Run 3. Situational awareness: This run aims to distract the subjects considerably from their main goal of operating the train. Thereby creating a lack of concentration on the main task, hence a poor situational awareness. This is accomplished by asking the subjects to do tasks that are completely different from their goal, to perform continuously an unrelated secondary task.

To keep the scenario simple and subjects distracted enough, the subjects are asked to play a casual game on a smartphone - the secondary task. The administrator monitors and reminds (if needed) to do both in parallel. To prevent subjects getting used to the secondary task, after a given time they are asked to switch to a more distracting secondary task, of similar nature. It is still a application on a smartphone, but significantly different than a casual game (identifying similarity of moving shapes).

Some additional details are given below:

- Secondary task 1: The subjects are asked to play a game on a smartphone for this simulation run while driving from the very beginning. A slight familiarity with the game is desired to not distract them completely.
- Secondary task 2: At a predefined point in the scenario (approximately the later 1/3rd of the scenario) they are instructed to switch to a different secondary task. It presents the subjects with a dynamic graphic and they need to determine the similarities, if present or not within a given time. Most subjects are unfamiliar with this task, and thus demands more attention than secondary task 1.

Run 4. Task Load: This scenario simulates a poor task load by increasing the number of tasks required needed to be performed by the subjects, in addition to the main goal of driving. These additional tasks are however related to the main task, thereby minimizing additional distraction. The subjects receive a brief explanation of these additional tasks before starting this run.

The first part is the subjects are asked to observe the 3D view and note the occurrence of some signals (Nf and marker boards). The second part is a small questionnaire taken from a good practice guide on cognitive and individual risk factors' [[Rail Safety and Standards Board \(RSSB\), 2008](#)]. These questions are aimed at helping the driver 'stay in the loop' of the driving task. They do not demand too much cognitive resource, and the information is very much related to the primary task nevertheless adds additional tasks to perform. Thus, these additional tasks are essentially an observation and response to questions on a paper form. Two paper-based question forms are given to the subjects to respond, explained as follows:

- Form 1. It involves observing the 3D view of the track from the train cabin. They are asked to note down some train location related data (LRBG and train distance) available on the DMI, as soon as they see a signal marked NF, and any other track-side information boards (except F signals).

- Form 2. Upon receiving cue from the administrator (in the final moments of the scenario run) the subjects start answering the following questions. The responses can be as time and or distance, both as shown on the DMI.
 - What color was the last signal?
 - What indication was the last signal/marker board?
 - What was the last (single) beep on the DMI (update of information) that I noticed? - What new information was displayed?
 - Where precisely am I on my route (location)? (Use distance information available on the DMI)
 - Where is my next signal? (At what distance/time)
 - When should I be slowing down?
 - Am I traveling at an appropriate speed regarding speed restrictions and external conditions?

Run 5. Time Load: In this scenario run, the train is programmed to start with a time delay of 2 minutes vs. the given timetable. The subject is told that his/her train has a starting delay, and is asked to try to complete the session according to the given timetable. Effectively they have 9 minutes for a scenario that they normally had to complete in 11 minutes. Given the fact that a tolerable delay is of 1 minute. Other than this time delay, this scenario does not impose any other conditions.

6.2.2.3 Explanation and basic training

The first step in the implementation of this protocol as an experimental campaign is a basic explanation and training session, as shown in the **Figure 6.2**. This is to explain to the subjects their main goals and performance objectives. These objectives will be later used for evaluating the performance. Since these are closely linked – the performance evaluation criteria and what the subjects know, thus this explanation session is a part of the experimental protocol.

It consisted of explanation on the following points:

- An explanation of the objectives of the campaign.
- ERTMS/ETCS signaling principles.
- To-the-point training of other signaling aspects, and driving a train under ETCS (DMI, procedures, etc.)

- A train driver's primary goals.
- Other specific explanation (messages, particular procedures, etc.)

Further, as explained to the subjects, this work considers the following set as a train driver's primary goals:

- Ensuring safety: it includes but is not limited to observe the speed limits on the DMI; observe all signals, marker boards, etc.
- Respecting standard operating procedures.
- Ensuring on-time service.

Essentially, for HRA objectives of present work, the aim is that they goals link to an HFE or HFE-like construct. The safety and service objectives of a train driver will be used to evaluated performance as explained later.

6.2.3 Output data sources and analysis

This section details the data that is obtained from the simulator-setup and the runs. It also describes how that data is used (score calculations, usage of questionnaires) to be usable later in this chapter.

6.2.3.1 Objective data: source and calculation of scores

A log of train driving data (ERTMS/ETCS cab-data): For each simulation run a raw data file called *EuroCab.log* is generated (a description is given in [section A.2.3](#)). As explained previously, an OPSIMUx simulates a DMI and a train cabin. Each OPSIMUx also saves this log file for each simulation run. Thus, raw data, specific to each train and each simulation run can be obtained from this file. This file contains, for a complete scenario data about some essential parameters. For a given data point (each is marked with the string "SPEED" for speed-related information, "RADIO" for radio-related information, and so on) the following information is given:

- **time** (in seconds) counts incrementally from the start of the scenario
- **distance** (in meters) is distance traveled by the train from the start of the scenario and 'front end location' an ETCS parameter used to locate the head of the train.
- **front end location** (in meters) it is the location of front end of train.

- **train speed** (in *km/h*), gives the actual speed of the train for a given time, followed by all of the ETCS speed curve speeds (more explanation in the next section).
- driver interaction (acknowledgment, data entry etc.)
- on-board and signaling data (automatic brakes, RBC messages, EOA location)

Thus, basic parameters of train position and speed are extracted from this file. It can also be used to extrapolate other parameters such as acceleration, driver reaction time, etc. To ease data collection a bash script was written, which asks for the subject and other identifying information and automates the data collection after each run. This log file is the raw objective data, saved specific to each simulation that will be run.

The main goals of a train driver are explained to the subjects in the explanation session as described in [section 6.2.2.3](#). These goals gives us objectives against which a human performance is evaluated. We define these criteria using a *score* value. There are two types of scores, a safety component and a service component. Both of these scores are calculated for each run for a given simulator session.

As explained previously [section 3.4.1.2](#), ETCS braking curves form a crucial part of ERTMS/ETCS functioning. Since, the curves are defined by the on-board system applying safety criteria (ATP functionality). Thus, the present work uses these braking curves as a baseline to evaluate the operational safety of a train. These braking curves are compared against a train's speed, to give an indication of unsafe or safe state of the train. The first objective criteria evaluate performance, against the safety objective of *observing the speed limits* for a train driver. This information in ERTMS is displayed on the DMI of the train driver [section 3.4.1.2](#). It is also explain to the subjects in [section 6.2.2.3](#) as one of the primary goals of a train driver. This goal or task of *driving under the assigned speed limit* forms an HFE, in a HRA context and also PRELUDE's interpretation.

The safety score. It represents an objective criteria safe operation: *remaining under the speed limit*. Owing to the amount of precise data available in the form of braking curves, a continuous and dynamic safety related data can be obtained, instead of simple over-speeding not over-speeding-based criteria.

The curves in [Figure 6.5](#) shows a representation of all of the ETCS braking curves and a train's speed (in blue). Similar to the explanation in [section 3.4.1.2](#), they are: the outermost curve in red: Emergency Brake Intervention (EBI) speed, in orange: Service Brake Intervention (SBI) speed, in yellow: Warning speed, green:

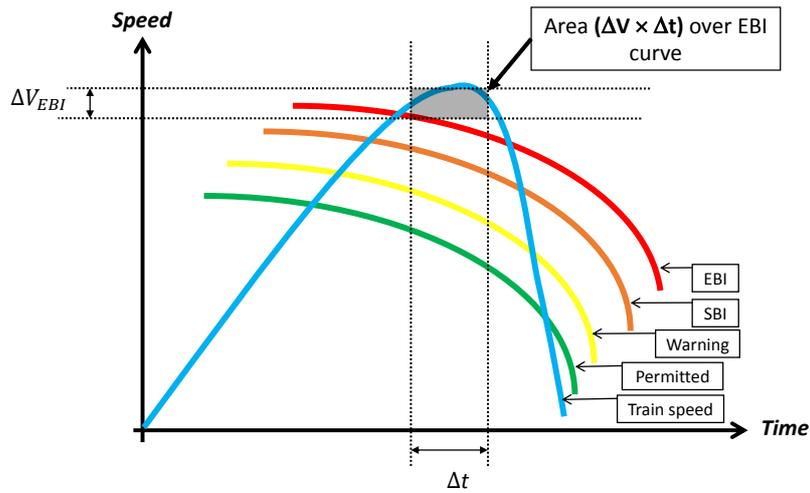


Figure 6.5 – Speed curve score calculation - in purple shaded area over EBI curve, that is the braking curve for the Warning component V_{EBI} .

Permitted speed. This speed data and train speed/braking data available in the log files as explained previously. To note that the x axis represents time, hence the speed-time curve for real data is composed of straight lines (where the slope gives the acceleration). Whereas a speed-distance curve is normally used and is in the form of a parabola. Nevertheless, these plots, they are given here only as an explanation aid. This combination of parameters, speed and time, is used to compute the score.

The objective of safety score is to determine a penalty for traveling at a higher speed than required, this penalty represents a bad performance related to the gap between the current speed and speed limit. Further, this penalty should differ between different braking curves. Because, quite evidently, not respecting permitted speed limit (P) is not the same as not respecting the emergency brake speed (EBI). This difference is accounted for using weighting coefficients. These weights generated from fixed ETCS values given in [Table 3.1](#), as explained below. Permitted speed limit is taken as a reference, and a normalized weight is calculated, which represents the degree of penalty greater than permitted speed. Similarly for each braking curve a parameter is calculate. The normalized weight for a given reference speed is calculated as follows:

$$wV_i = \frac{wV_i}{\sum wV_i}, \text{ where } wV_i = \frac{dV_i}{dV_P} \quad (6.1)$$

where i is the reference speed that is $V_P, V_W, V_{SBI}, V_{EBI}$; dV_i is the maximum values of fixed speed difference taken from the reference [Table 3.1](#); wV_i is the normalized weight for a given speed i .

To note that, we count having a speed greater than P permitted speed as

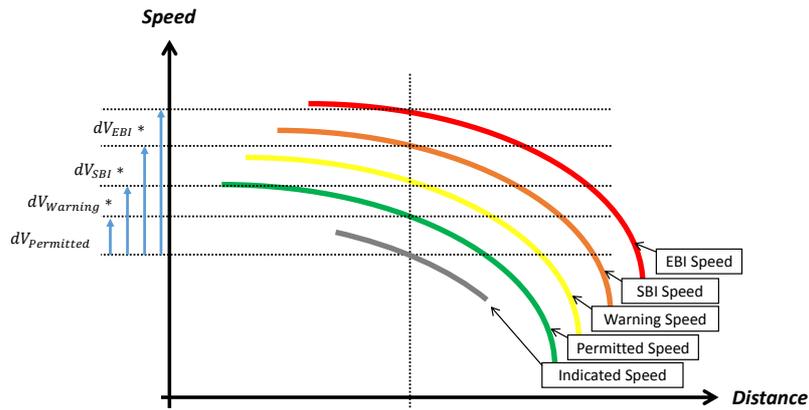


Figure 6.6 – Modified dV parameters to take into account speed difference from indicated speed for score calculation

Reference speed	dV_{i^*}	normalized weight wV_i
Permitted: P	5	0.1
Warning: W	10	0.2
Service Brake Intervention: SBI	15	0.3
Emergency Brake Intervention: EBI	20	0.4

Table 6.1 – Normalized weights for ETCS reference speeds

a minimum penalty. That is indicated speed I is the maximum allowed speed, which incurs zero penalty. However, the fixed values in [Table 3.1](#) are counted from P permitted speed, *not* indicated speed. To account for this difference, dV_W , dV_{SBI} , and dV_{EBI} are modified as $dV_{W^*} = dV_W + dV_P$, $dV_{SBI^*} = dV_{SBI} + dV_P$, and $dV_{EBI^*} = dV_{EBI} + dV_P$. Thus, all the reference speeds are measured from indicated speed, rather than permitted speed. These new values are shown in [Figure 6.6](#), and the actual values obtained are given in [Table 6.1](#).

Once weights are obtained, the final can be calculated. Thus, safety score is calculated as follows: for a given time interval Δt , the positive difference ΔV_i between the reference speed (immediately inferior braking curve) is calculated. This over-speed value (ΔV_i) is multiplied with the time this over-speed was observed for (i.e. Δt). This gives us the penalty for corresponding reference speed i . It is then multiplied by the weight from [Table 6.1](#) for the concerned reference speed (immediately inferior braking curve). This value is summed over all of the duration of a given scenario giving the final safety score, it is given by the following equation:

$$SS = \sum wV_i \times \int_{j=0}^T (\Delta V_{ij} \times \Delta t_j) \quad (6.2)$$

Where, SS is the final safety score for a given scenario of total duration T ; wV_i

is the normalized weighing factors associated with the i th ($i = P, W, SBI, EBI$, in order permitted, warning, SBI and EBI) braking curve as obtained previously; ΔV_{ij} is the non-negative speed difference between the i th braking curve speed and train speed, for a given time period Δt_j .

In [Figure 6.5](#) for a given time Δt and for a given over-speed against EBI speed that is ΔV_{EBI} , this gives the shaded purple area as the area on the graph, which when multiplied gives the score for EBI over-speed. It is then multiplied by the penalty $wV_{EBI} = 0.4$ as stated in [Table 6.1](#), to give the safety score component of the EBI speed. Similarly for all other reference brake curves and corresponding Δt . To note that, for this Δt , additionally there will be a *SBI* component, a *warning* component and a *permitted* component - since for this Δt the train is over-speeding against all these reference speeds.

Thus, the higher the Safety Score the worse over-speeding is, thus the unsafe a train's operation becomes. This is hence a cumulative measure, over the whole scenario of an unsafe behavior by the train driver by over-speeding. Its unit is meters (speed \times time), what we will refer in the following sections as *over-travel distance* or simply safety score.

The time score. It aims to capture the service component of a given simulation run. To establish baseline time, two experienced subjects (which did not take part in the experimental campaign) run the scenario in normal conditions. A buffer of 60 seconds is added to this ideal time to give a reference time. The timetable, thus created, is given to all of the subjects (a paper copy for reference) for all of the sessions. If a delay of more than the reference time is incurred, it is interpreted as non-accomplishment of service goal.

The time score is simply the time delay vs. the reference time (in seconds) that a train has when arriving at the end of a given run. If the train arrives on or before this reference time, time score is considered 0. If there is delay, the time score is calculated as follows: $max \Delta t$ where Δt is the positive difference between total time allocated (in the timetable) vs. the actual time taken.

A python script is created which takes as input data the *Eurocab.log* file, performs a regular expression text analysis to extract the relevant data (speeds: train speed, W, P, SBI, EBI , train position, start time, end time). And outputs the safety score and time score for each subject's each run. An example of the data in this log file and the python script used are given in annex [section A.2.3](#).

6.2.3.2 Subjective data: source

Subjective data, for present study refers to a self-assessment, using simple questionnaires and a standardized techniques. There are numerous methods to determine different aspects of human cognition from subjective techniques. Two subjective questionnaires are used, NASA TLX (Task load index) and a PSF-subjective questionnaire. Both of these questionnaires follow closely the objectives of an HRA study as is the objective of this work.

NASA Task Load Index: As given in [section A.2.1](#) NASA TLX provides subjective workload rating. TLX provides an adequate framework to obtain a general and standardized indication of the perception of the subject. Furthermore, TLX has sub-scales which can be related to PSFs. Hence it provides method to have a simple yet comprehensive data collection. Some of TLX's limits such as off-line administration (after the task has been completed) and time needed to complete and analyze the test are not major hindrance for present work. A windows desktop tool [[Sharek, 2009](#)] was used to administer NASA TLX, also seen in [Figure 6.7](#). As a reminder the list of TLX sub-scales on which subjects are questioned are given below:

- Mental demand
- Physical demand
- Temporal demand
- Effort
- Frustration
- Performance

General details and definitions are given in [section A.2.1](#). As a reminder we give the equation using which TLX scores are calculated as follows

$$TLX \text{ Score} = \frac{1}{15} \sum_{i=1}^6 D_i \times C_i \quad (6.3)$$

where: *TLX score* is the global workload score, D_i is the raw sub-scale rating for the descriptor i , and C_i is the number of times a descriptor was chosen in the pairwise comparisons. The sub-scale scale score or the *TLX Scores* can be used when evaluating or comparing a task. The output are TLX's individual sub-scale values, respective weights (after the pairwise comparison), and finally the TLX rating (or TLX global score). These outputs will be used in the analysis of subjective data, as described later in this chapter.

PSF-subjective questionnaire: A simple feedback is used to complement the subjective data collection process. We are concern with the most critical PSFs which are taken from the PRELUDE methodology [Table 5.1](#).

In this questionnaire, the subjects rate PSFs in terms of their (perceived) influence on performance, for a given run. To respond they can select any one of the four rating levels (*Good, Nominal, Poor, Not sure*). However, since in the experimentation the subjects are not necessarily experts, they were provided with simpler definitions and vulgarized examples. The questionnaires were given to the subjects using google forms. It is also provided with this manuscript in [section A.2.2](#). This questionnaire included in total 8 questions. The first question asked if they felt the situations or conditions were real, and the second inquired on the general perceived difficulty level. The rest of the questions (for the 7 PSFs in total) were about the PSFs. The subjects respond by selecting one our of four possible states, based on perceived influence on their performance for a given run.

A third a set of pre-simulation and post simulation questionnaire as given in [section A.2.2](#) and [section A.2.2](#). They aim to obtain an indication of the state of the subjects, pre and post simulation.

6.2.4 Experimental campaign and preliminary discussion of the collected data

This campaign implements the previously proposed protocol. The later sections aim to demonstrate what data be used (objective, subjective), and how can it be used. If the need be the campaign can be easily be carried out in an industrial setting with real train drivers (operators). This works use of standardized signaling context, data sources and tools will allow for easy implementation of this protocol. For these reasons, we believe a demonstration with university students is adequate enough for a first implementation of this work. Further, in such a case a reasonable number of subjects are sufficient, we believe 13 is such a sufficient number. Thus, a total of 13 volunteer subjects (university graduate and undergraduate students) participated in the experimental campaign. As stated before these are subjects, not domain experts.

An overview of the experimental campaign's undertaken is given in [Figure 6.7](#). It also include some pictures of the subjects performing different activities as part of the experimental campaign. Each of the subjects were given an explanation on the objectives of this experimentation, and signed a consent form accepting to take part in the campaign. The sessions were mostly performed in the afternoon, with each subject taking a complete session in one sitting. A session for a given subject

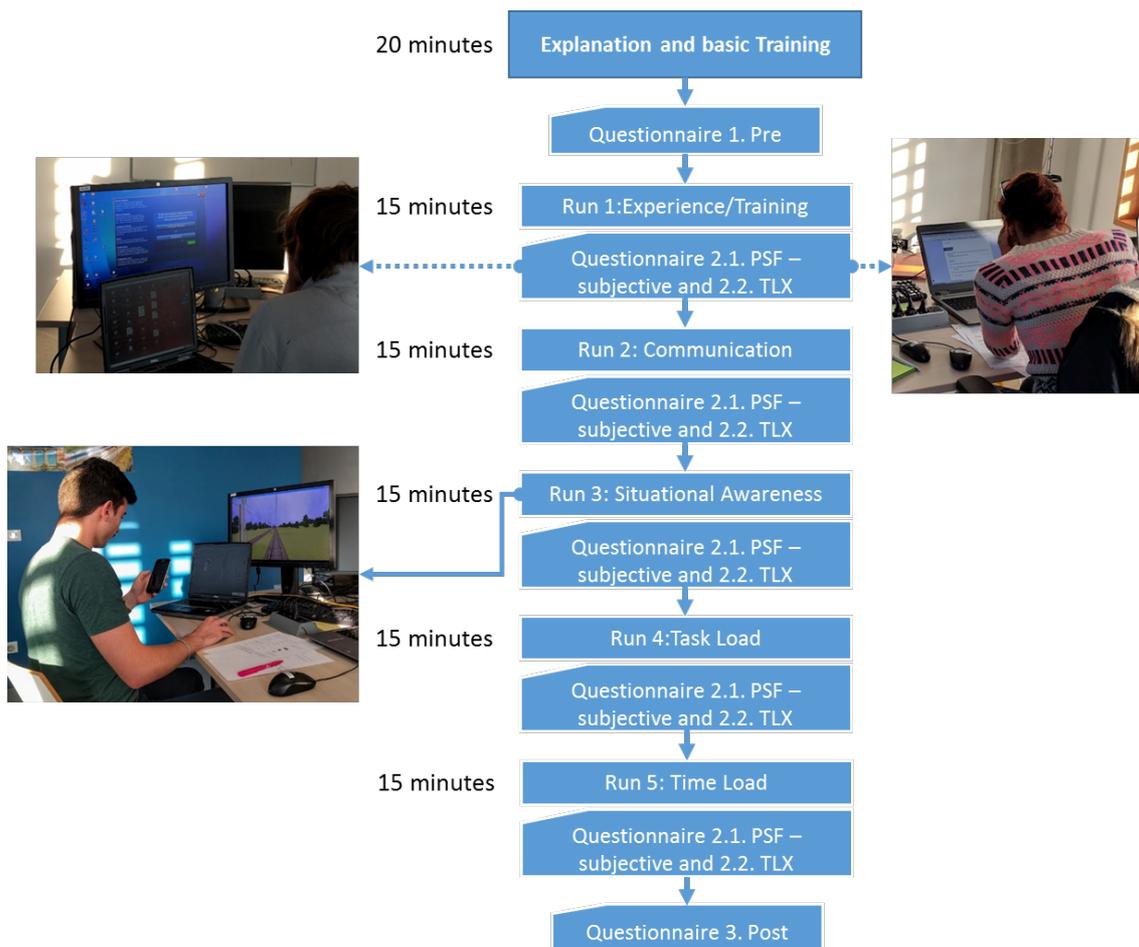


Figure 6.7 – The experimental campaign with some photos of subjects during the course of a session

lasted about 2 hours.

Objective data was collected as explained in [section 6.2.3.1](#). The subjective questionnaires were administered using google forms. And the TLX using a desktop software application as discussed previously. Thus, the data collected is as follows:

- Objective data
 1. *Eurocab.log*: Train speed and ETCS braking curve data (for all runs)
 2. Additional run-specific data
 - Run 4. Task load ([section 6.2.2.2](#))
- Subjective data
 1. NASA TLX responses (for all runs)
 2. Subjective data on perceived PSF's states – PSFs (for all runs)
 3. Additional questionnaires

- Pre-questionnaire (once per-subject)
- Post-questionnaire (once per-subject)

Once the experimental campaign was carried out, the next section briefly presents and discusses the data obtained. This data will be analyzed and used in the later sections. This section aims to present a preliminary discussion on the data that was collected.

6.2.4.1 Objective data

Following the description and computation of safety score in [section 6.2.3.1](#), a script (provided with this manuscript in [section A.2.3](#)), is used to extract and compute safety score. A safety score is generated for all of the simulator runs performed by a subject. 13 subjects and 5 runs each, gives us 65 (13×5) sets of safety and time scores.

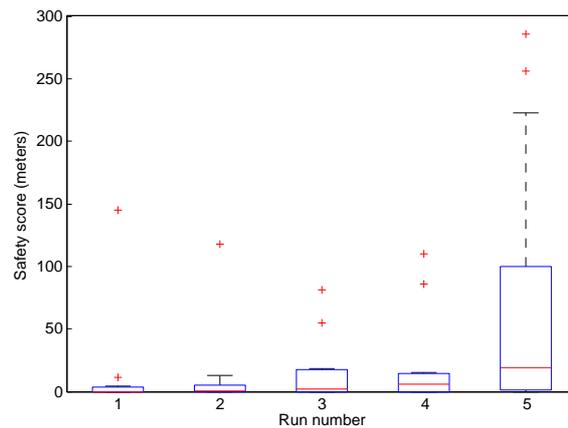


Figure 6.8 – Box plot of the *Safety score* of all the subjects for each run

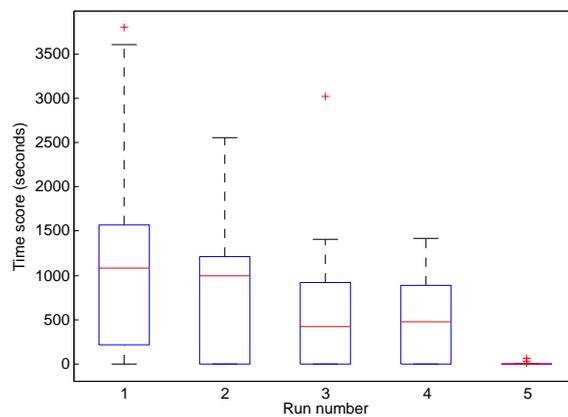


Figure 6.9 – Box plot of the *Time score* of all the subjects for each run

The plots in [Figure 6.8](#) and [Figure 6.9](#), give the these two scores for each subject, for all the runs. These raw scores are presented as a box plot, it includes: the median, the 25th percentile, and the 75th percentile.

We briefly comment on these raw scores as follows. For safety score [Figure 6.8](#) the minimum value is zero. The median values show a increase along the runs, and particularly **for run 5 see a significant increase**. Similarly the time scores are shown in [Figure 6.9](#): as can be seen when median decreases between run 1 and run 3, then a slight decrease for run 4, and then drops down to almost zero (for almost all of the subjects) for run 5. The increase and decrease, shows the need for further analysis of results.

6.2.4.2 Subjective data

NASA TLX global and sub-scale scores: We start by presenting the average TLX scores and ratings for all the subjects and all the runs. As described in [section A.2.1](#), NASA TLX aims to obtain a subjective estimation of workload. As presented in [section 6.2.4](#) the subjects respond to a TLX questionnaire and a PSF subjective [Equation 6.2.3.2](#), after each run.

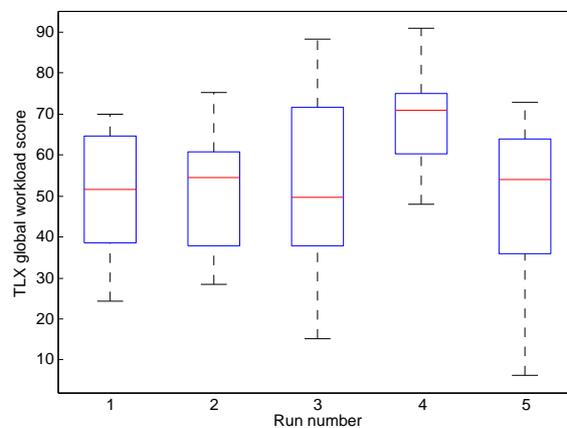


Figure 6.10 – Box plot of the TLX global score

In [Figure 6.10](#) we see the TLX global workload score for each run, presented in the form of a box plot. Some comments on the global workload score are as follows:

- It can be seen that, for all the runs 1, 2, 3, 4 a TLX global rating of more than 50, which signifies a moderate workload.
- The Run 4 is where the highest, and most consistent (smallest box height) TLX score is obtained. This is expected, because of the additional tasks required in Run 4 ([section 6.2.2.2](#)) and the explicit focus of TLX on workload. This,

confirms the hypothesis that increased task load was indeed 'simulated' for the subjects, and it was sensed as an increased subjective workload by the subjects vs. other runs.

Subsequently, [Figure 6.11](#) presents the (average for all subjects) scores for sub-scales, where the top continuous line shows the raw rating value (the scale for a given parameter) and weighted value a given parameter. The weighted values are obtained after a multiplication and normalization, (see [Equation 6.3](#)). The scale ratings are values on a numeric scale between 1 to 100. Some comments on average vales of these rating scales are given below:

- Overall for all the subjects **relatively similarly evolving ratings were obtained for mental, temporal and effort** (see top three images, [Figure 6.11](#)), an increase after Run 1, 2 and then 3, with the maximum for Run 4, and finally reducing for Run 5.
- The values of average **ratings for mental, temporal and effort TLX were higher** than other ratings: that is physical, performance, and frustration. ([Figure 6.11](#) bottom three plots). This indicates this experimentation protocol was biased to mental and temporal tasks. This is expected, since the scenario was non-physical (sitting down and using a mouse and keyboard to drive the train).
- For Run 1, subjects rated their performance the highest. With a steady decrease thereafter. Indicating more time spent on the simulator, made them learn more on their performance, even though the safety score remain either unchanged or decreased ([Figure 6.8](#)).
- Run 3 was the second highest in terms of mental demand, temporal and effort.
- Run 4 higher (vs. other runs) values of rating scales for mental, temporal, effort, and physical were obtained. They were also more frustrated although not by a significant margin.
- **Run 5 saw a decrease** compared to previous run, for almost all the ratings. However, **safety score show a sharp increase**, as can be observed in [Figure 6.8](#). Indicating that **even though workload was low, the subjects performed poorly in terms of safety**, prioritizing service (almost zero time score [Figure 6.9](#)).

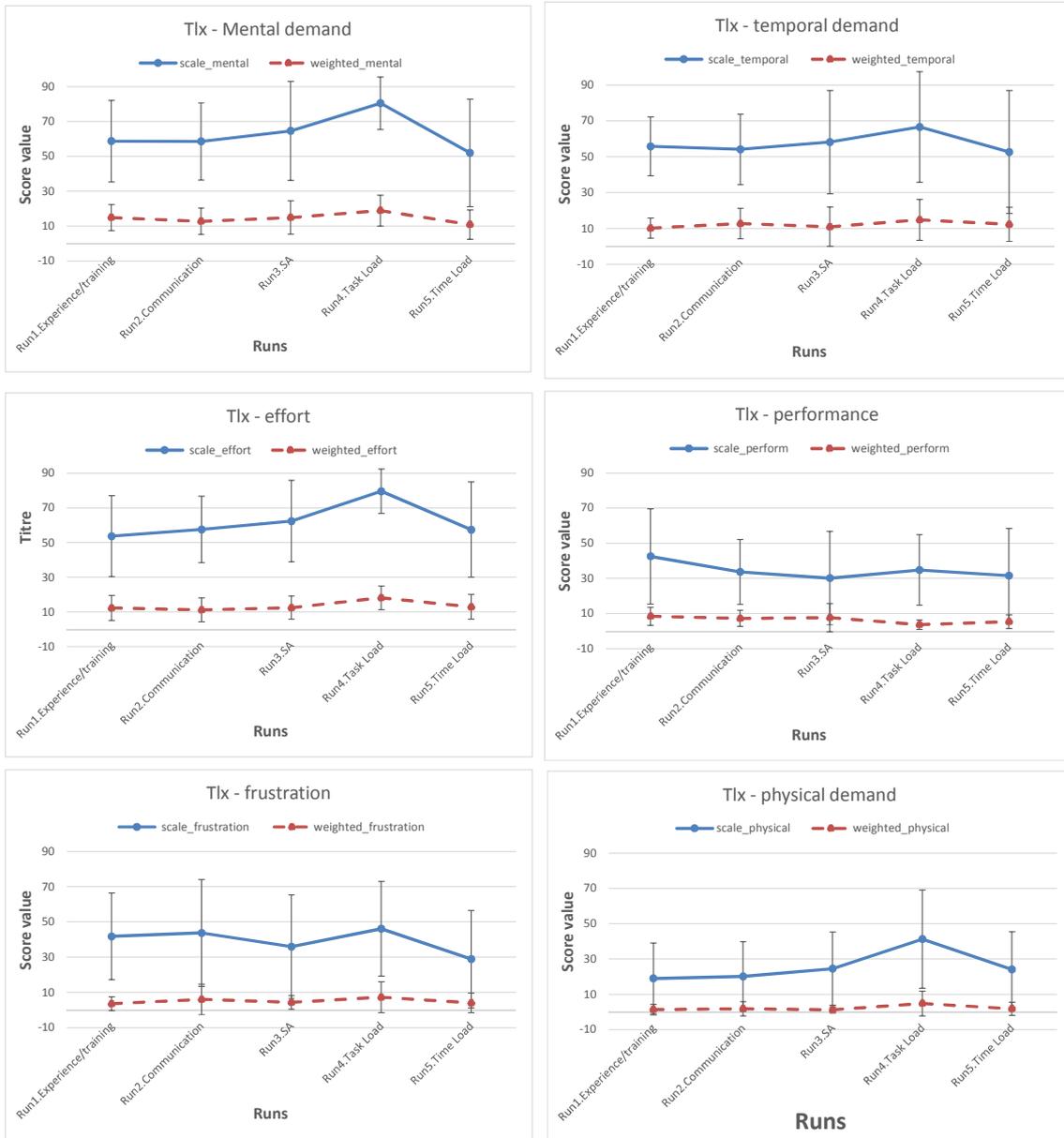


Figure 6.11 – Average TLX sub-scale scores for all the parameters of TLX: ratings and weighted values

PSF subjective data: The responses to PSF subjective questionnaires are described in this section. Since responses are obtained as discrete values they are presented as a count (number of a certain responses). Further, as described in [Equation 6.2.3.2](#) this work is interested in the degrading conditions, i.e. *poor* the responses indicating a degraded state of PSFs. Thus, the responses given as *poor* (red) and *I am not sure* (gray) are counted for a given run and shown as count plots. Here, the responses for each run are analyzed and compared against other runs. These results are discussed as follows:

- In run 1, **majority of the responses were indeed poor for Training and Experience**. Over 60% (8/13) of the subjects said they had poor experience. Over 40% felt the training was poor, and some also responded to Time Load being poor. Both of these results validate that run 1, did indeed, from subject's perception conditions of a poor Experience/Training.
- In run 2, there were 4 out of 13 responses to communication being poor. Although not in majority, compared to other questions for this run (other PSFs) **relatively highest number of subjects responded as communication being poor**. Nevertheless, these results indicate that subjects did not perceive, the communication tasks as degrading their performance.
- For run 3 it can be seen that **most of the subjects responded that their Situational Awareness being poor**, more than 60%. However, most of them also felt (1) that the conditions/situations in this simulation were not real, and (2) they perceived it to be difficult than the previous runs. This for one validates that in this run, majority of the subjects felt their SA was poor, and had an negative effect on their performance.
- For run 4 **most subjects perceived situational awareness (over 50%), Task load and Time Load**, with no other PSF being reported to be poor. This run was designed to avoid poor Situational Awareness perception, the additional tasks (refer [section 6.2.2.2](#)) were related to the main driving task. However these results, indicate that the subjects felt distracted, and the additional tasks increased the global workload (task and time both). Furthermore, most of the subjects (over 60%) felt this was a difficult scenario.
- In run 5, for all the questions almost all the subjects did not perceive any PSFs in a poor state degrading their performance. Similar, observations were made while analyzing TLX results indicating the 2 minute delay was not understood by the subjects as an significantly increased time load. It can also

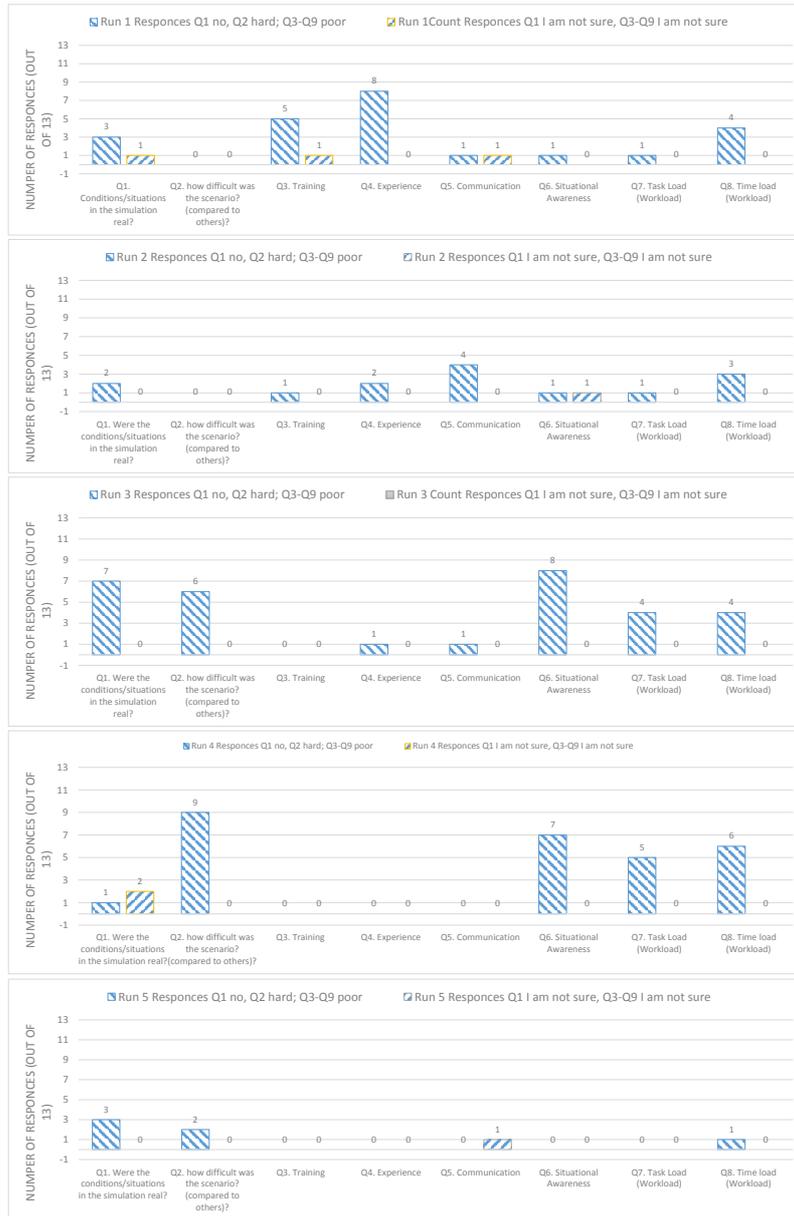


Figure 6.12 – The number of responses to PSF subjective questionnaires, “poor” and “I am not sure”, for all the runs

be attributed to the fact that there was no immediate penalty imposed time delay, apart from the instructions given at the start.

- In general for runs 3, 4, and 5 as the subjects gained some experience, the **Training and Experience saw reduction in number of poor responses.** The focus then shifting towards other PSFs. And vice versa in the case of SA, Time load and Task Load for first two runs.

For majority of the runs (run 1, 3 and 4) - it was observed that majority of the subjects perceived the degraded states of PSFs as expected (section 6.2.2.2). For other runs, either subjects did not understand or agree with the definition of the PSFs (run 4, 5); or the conditions simulated were not degraded enough for the subjects to perceive as degrading their performance (Run 5, 2). This concludes the experimental protocol description, the next section details how the data is used in the PRELUDE methodology.

6.3 Using data obtained from experimental protocol in PRELUDE

As discussed previously we use the data collected from the experimental protocol in PRELUDE. This section details this proposition. We start with performing a pre-analysis of data and then subsequently using them in the quantitative and qualitative part of PRELUDE. An overview of this usage was introduced before Figure 6.1. A more detailed version of both usages will be given while explaining each proposition.

6.3.1 Pre-analysis: A classification of subjects

Looking at the data collected in the previous section, there is a need to pre-analyze the scores. This pre-analysis is performed because of the following reasons: (1) A safety-focused model like HRA should account for the 'worst' in terms of the performance, such that a lower bound approach towards operational safety can be taken; and (2) to better understand the reasons (or factors) why the performance was inadequate, mainly the individual differences and reasons.

We are interested in a subject's personal performance, thus scores for each of the run for a subject are averaged. Thus, what we get is an average safety score value for a subject's all of the runs. Instead of analyzing one by one, we chose to analyze these differences in terms of small groups. Thus, a classification of subjects based

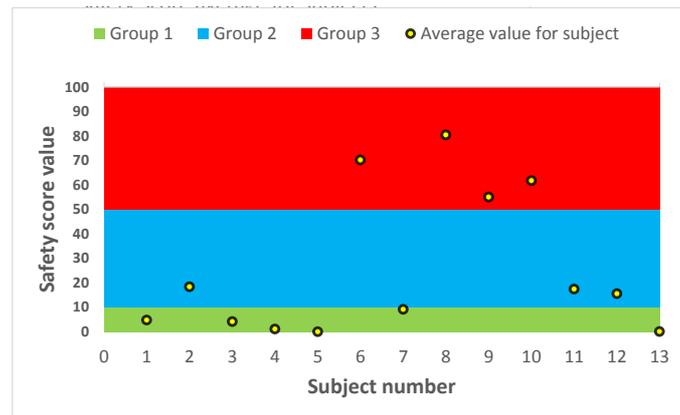


Figure 6.13 – Classification of subjects into groups based on their average scores

on their safety score is presented in this section. This classification is a pre-analysis which aimed at considering the two points previously discussed.

Since we have 13 subjects, we make a choice of having three groups. This choice is reasonable, since it will represent the trio of high, medium and low. In **Figure 6.13** the data points shows the average values of safety scores for each subject over all the runs performed by that subject. Such a visual representation allows distinguishing the inter-subject differences, rather than inter-run which was discussed in the previous section.

For the data points in **Figure 6.13** we are interested in the vertical separation (safety score values) between different subjects. It can be seen in that for a safety score of more than 50, an upper cut-off be selected for the *high safety score group*. Since there is significant separation between subjects with safety score below that threshold.

Further, a value of below 10 can be assigned another *low safety score group*. That leaves us with a value of safety score for a subject between 10 and 50, to be classified as group with *medium safety score*. After the identification of thresholds the subjects are assigned into groups. These groups are defined in terms of subject-wise average safety score, as follows:

- *Group 1*: safety score value less than 10;
- *Group 2*: between 10 to 50;
- *Group 3*: more than 50;

Thus, in **Figure 6.13**, a color code is overlay onto to show the subjects in different groups. Green shows the subjects of group 1, blue for subjects in group 2, and red for group 3. Subject number 08, 09, 10, 12, and 06 have, for each of them a safety score less than 10. Subject number 01, 02 and 11 have a safety

score between 10 and 50 that is group 2. And lastly subjects number 04, 03, 07, and 05 have a safety score more than 50 thus they are classified in group 3. To note that statistical methods (\sqrt{n}) can be used to determine the number of groups, here however we have used such empirical reasoning since it is sufficient for the data that we have.

Having classified the subjects into groups, we draw box plots again, this time a different plot for all the subjects in a group. [Figure 6.14](#) and [Figure 6.15](#) shows the safety score and time score respectively for the three groups. As it can be seen that this grouping allow us to work with relatively coherent data. Compared to the scores of all the subjects ([Figure 6.8](#) and [Figure 6.9](#)) the results after grouping are much more coherent. That is the differences are much more visible.

Furthermore, group 3 has safety score which are the highest (worst performance) among the three groups. We consider that the worst case (data), that is the lowest performing subjects, are the most representative of the worst safety cases. Thus, they will be used in the next sections to model human reliability, and inter-group comparisons will be made to analyze other factors for their low performance.

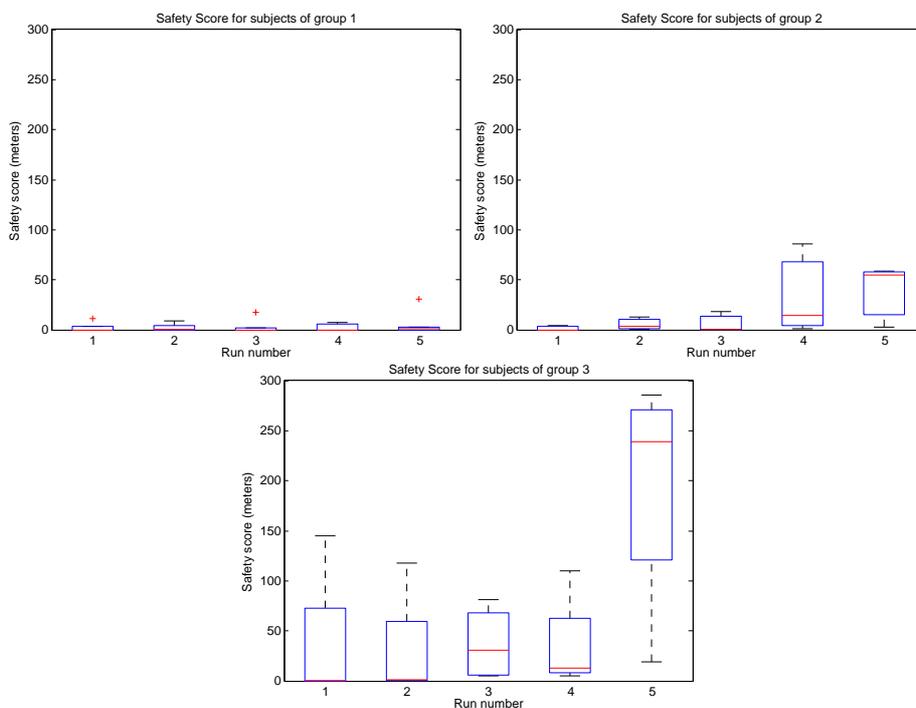


Figure 6.14 – Box plot of safety scores for subjects in group 1 (top left), group 1 (top right) group 3 (bottom)

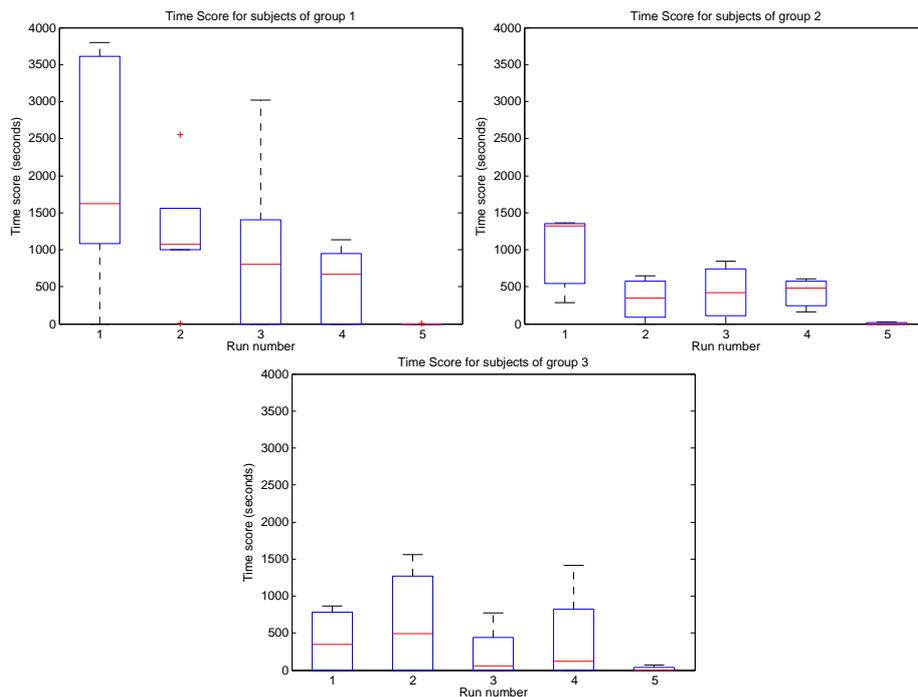


Figure 6.15 – Box plot of time scores for subjects in group 1 (top left), group 1 (top right) group 3 (bottom)

6.3.2 Objective data usage: combination with expert data – input to PRELUDE Quantitative part

This subsection details how the objective data obtained from the experimental protocol is used in the PRELUDE’s quantitative part. As shown in [Figure 6.16](#) after the pre-analysis presented in previous section, there are four steps, starting from adapting the data to the final modeling in the VBS. With the exception of the first step, other steps are proposed such that they are similar to PRELUDE’s quantitative methodology, to make it easier to use this data. Thus, the main usage presented in this section is to define the relations between the PSFs and HFEs. This is similar to the questions asked from the experts in PRELUDE’s quantitative section.

As defined in the protocol we have an objective criteria (the tasks that a train driver has to do, time delay to respect) and a numerical indicator on that criteria (safety score, time score). In addition we also have data on the PSFs’ states: as a run was created (refer [section 6.2.2.2](#)) that is for a given run a specific PSF is *poor*. This allows us to obtain a conditional piece of information, notably on the state of a PSF and the numerical indicator on the objective criteria. In previous chapter conditional data on a state of PSF to a state of HFE with a probability value was elicited from the experts.

Now as with expert data, such experimental data can only be used, with exten-

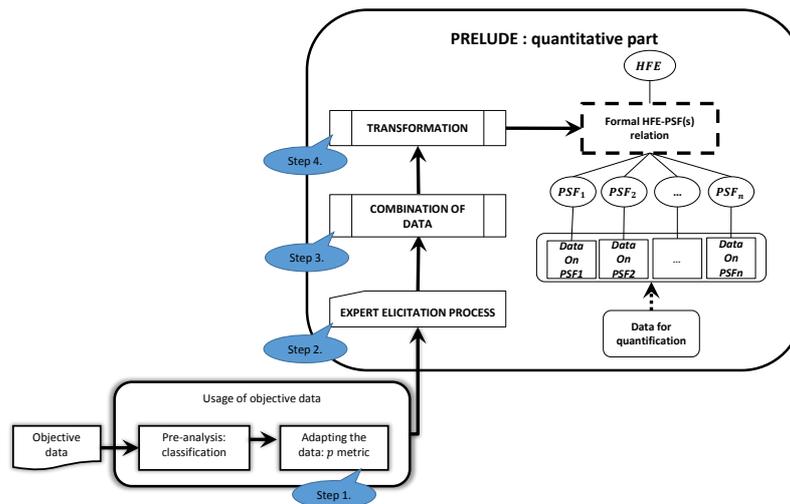


Figure 6.16 – Overview of the usage of objective data from simulator experimentation in the PRELUDE's quantitative part

sive data collection - in terms of the frequency (number of subjects) and scenarios (HFEs, PSFs/conditions, etc.). However, such an activity poses considerable cost and time requirements. A relatively simpler approach is to complement expert data. Thus we propose to use this objective data to complement expert data. Further, to demonstrate the proposition we re-use expert data from PRELUDE's case study [section 5.2](#). To note that, only safety score is used to demonstrate this proposition. This objective data is treated similar to data from another expert. That is we combine this data with other expert data before building the VBS model. This approach is also similar to PRELUDE [section 5.1.2.2](#), to combine data. That is, data from different sources are combined at the first level, and then a second combination builds the model.

The following steps, also seen in [Figure 6.16](#), give a detailed explanation on how this data is *adapted*, *combined* and *transformed* to build the model:

Step 1: Adapting experimental data: calculation of p metric for objective data

Step 2: Selection (or elicitation) of expert data

Step 3: Selection of combination rule(s) and combination

Step 4: Transformation to build VBS model(s)

Step 1. Adapting the data: calculation of p metric from objective data: Since in the context of this work and as presented in the previous chapter, PRELUDE methodology works with probabilistic data. That is probability values are elicited from experts. Thus, the objective data needs to be transformed to probability value.

This can be a simple form, e.g. total number of errors committed divided by total possibilities, or relatively complex depending upon the data available. To do that, we introduce a p metric, p for probability. This value aims to obtain a probabilistic value from the objective data, the scores in this case.

As a recall the safety score is defined as (refer [section 6.2.3.1](#)) *over-travel distance*, which is normalized distance traveled by the train while over-speeding. Thus, from the standard notion of an error probability, we have the following: $p(\text{error}) = \frac{\text{total number errors}}{\text{total opportunities for error}}$. This, when applied to a train driver over-speeding is interpreted as follows: total opportunities is the total length of the track section (in meters), and total number of errors is the distance travelled while over speeding (or the over-travel distance) i.e. corresponding safety score (in meters [Equation 6.2](#)). This gives us the equation to compute the probability metric of safety score p_{SS} as follows:

$$p_{SS} = \frac{\text{Safety Score}}{\text{Total Distance}}$$

This equation gives an adaption of safety score as a probability value. Once adapted, this objective data can now be combined with expert data.

Step 2. Selection (or elicitation) of experts data: As discussed before, we plan to combine expert data with objective data from the protocol. Hence, once the simulator data is appropriated, expert data needs to be obtained. This data can come from a pre-existing expert elicitation process (discussed below), or another elicitation process can be carried out. For demonstration of this step, we will employ the data already elicited from experts for the case study of PRELUDE methodology [section 5.1.2.2](#). An evident but important condition here is that both of these data need to be about the same entities, i.e. the HFE and PSF. One of the HFEs used in the PRELUDE case study matches the criteria behind safety score. To note that PRELUDE expert elicitation was carried out for the case study, and not for an ERTMS context, nevertheless for present demonstration this difference is acceptable. In [Table 5.7](#) the HFE3 is defined as:

- **HFE.** (HFE3 in PRELUDE's case study) Not reducing speed in time.

Definition: *National regulations* "any agent, regardless of his/her roles, should respect the concerning signals . . .".

"The driver shall endeavour to recognize the signs (signals) as far as possible and do not lose interest in their observation as (long as) it (train) has not crossed them." Further, ". . . driver should identify the reference (point) to initiate the braking and to reduce the speed."

In the case of cab-diving these reference points are identified on the cab driver's on-board display (section 6.2.3.1), and more information is available using line-side signals.

As specified previously in section 6.3.1, group 3 has the highest safety score (worst safety performance) of the three groups. And this is used to consider a lower bound on safety considerations, that is the model considers the worst performance cases. Thus, we consider the p metric for group 3 to combine with expert data. Here, in Table 6.2 the p metric of group 3: p_{SS}^{G3} is given. It is the average value for a given run of all the subjects in the group 3. The data from experts obtained in ?? is also given. To keep interpretations and combination straightforward the intra-source data is pre-combined. That is, data from different experts (about a question), and scores for different subjects (for a given run) are combined amongst themselves using a simple average combination. The pre-combined expert data in PRELUDE is given in ?? for questions 3.1, 3.2 and 3.3. We recall and present both of these data in Table 6.2 along with the questions and scenario run definitions.

Table 6.2 – Expert combined data and data from experimentation

Question for experts	Combined probabilistic response (average)	Scenario run	objective data (average for p_{SS}^{G3})
Given the occurrence of a poor level of <i>Situational Awareness</i> , what do you think about HFE being true?	0.04	The p metric of safety score for Run 3. which creates condition of <i>Situational Awareness</i> poor	0.00195
Given the occurrence of a poor level of <i>Experience</i> , what do you think about HFE2 being true?	0.04	The p metric of safety score for Run1. which creates condition of <i>Experience/Training</i> poor	0.0019
Given the occurrence of a poor level of <i>Time Load</i> , what do you think about HFE being true?	0.04	The p metric of safety score for Run 5. which creates condition of <i>Time load</i> poor	0.0103
Given the occurrence of a <i>nominal</i> level of <i>all the PSFs</i> what do you think about HFE being false?	0.95	Not available (considered same as expert data)	0.95

Step 3. Selection of combination rules and combination (PRELUDE quantitative): Once the data is obtained from both the sources, it needs to be combined. Again a similar approach to PRELUDE (section 5.1.2.2) is proposed, where data from different experts (sources) was combined as a first step using different combination rules. To that extent, it can be considered that expert and simulator data for each question/run in Table 6.2 is independent. Thus, we can employ the combination rules as used previously in section 5.1.2.2.

In order to combine with the BFT-based combination rules (Dempster's and Yager's rule) this data needs to be modeled in a BFT format. As reminder, this is

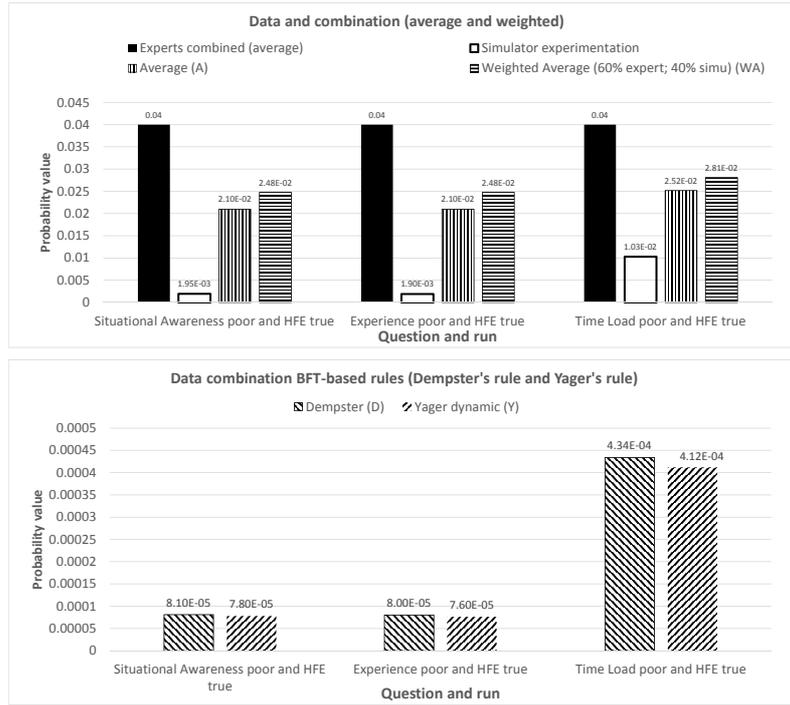


Figure 6.17 – Expert data and simulator data (p_{SS}^G) value, and their combination using: average, weighted average (top), Dempster's and Yager's rule (bottom), line plot shows the conflict value $(1 - k)$ on the secondary x axis.

done as follows:

For a proposition (what the question aims to measure) say 'X is exactly x and only x ' an expert's belief is represented by a *BPA*. The value of this *BPA*, say b , is a quantitative expert belief (a subjective probability) on the said proposition. Each expert's response is then modeled as a complementary belief structure. This goes to state that, for each expert, belief about the value of X being x is b and exactly b . Therefore, the belief of $X = \{\bar{x}\}$ is $1 - b$. This is then modeled as two focal sets with the associated *BPA* values. The belief structure in 6.4 gives the considered representation of expert data.

$$\begin{aligned}
 m(\{x\}) &= b \\
 m(\{\bar{x}\}) &= 1 - b \\
 m(\Omega_X) &= m(\{x, \bar{x}\}) = 0
 \end{aligned} \tag{6.4}$$

These considerations also hold for experimental data from the definition of the safety score, where it gives us an information about a subject over-speeding. Although here it is an experimentally obtained probability rather than an expert's belief. Once both are modeled as given in 6.4, they are combined and compared.

The Figure 6.17 shows the data from both the sources thus obtained and the results of combining them using different combination rules. Average and

weighted average combination (considering a hypothetical 60% weight to expert data and 40% for data from simulators) give a value of similar (10^{-2}). However, for combination with BFT-based rules, for low conflict value ($1 - k$ - which represents conflict - higher the value the higher the conflict), the case of *Time Load and HFE true*, the combined data obtained is higher. Here, it can be seen that Dempster's rule is reinforcing the consensus [refer: PRELUDE expert data hypothesis]) when the conflict is higher. As seen for 'Time load and HFE true', where the data from experts and simulator is of same order (10^{-2} ; [Figure 6.17](#) top left, first two bar plots) the combined data obtained is around 10^{-4} . Whereas, for the other two questions where expert (10^{-2}) and simulator (10^{-3}) data have high conflict, this leads to the combined data being 10^{-5} . Other combination rules (Dubois & Prade; Inagaki, not presented here) give similar (10^{-4}) results. Furthermore, we can see the expert data is of the order 10^{-2} and data from for group 3 is 10^{-2} or 10^{-3} , for group 2 it is less (less safety score than group 3, thus less p value also), and less further for group 1 (lowest safety scores). This shows that the experts' opinions are closer to the worst performance cases. This is possibly an indication that experts are overly conservative in their estimations and provide higher probabilities than what the empirical data shows.

Nevertheless, a limitation of a simulator experimentation is that for a given PSF, all scenarios with conditions of its degraded state are too impractical to simulate. For example a *poor* situational awareness can be distraction, mental pre-occupation with another task, lack of a mental picture of the system, etc. Whereas it is relatively easier for experts to consider such large definitions, based on their experiences. Thus, a trade-off between these two sources can be made based on what kind of HFEs and PSFs are involved. Further, if they are to be combined care should be taken where the conflict is high, and further investigation is needed before selecting the 'correct' choice of combination method. PRELUDE leaves this choice open to the analyst.

Step 4. Transformation (PRELUDE quantitative): Once combined data is obtained, it is input to the PRELUDE methodology as shown in [Figure 6.16](#). Since, the current approach is to treat expert and data from the experimentation at the same level, after combination, it follows similar transformation as data as in PRELUDE methodology [section 5.1.2.3](#). As a reminder this steps involves a vacuous extension followed by a combination of the questions using Dempster's rule. This then gives us a final configuration belief structure of the VBS.

To complete the analysis, we present the quantification of the HFE using generated VBS model. A different VBS model is generated depending on the

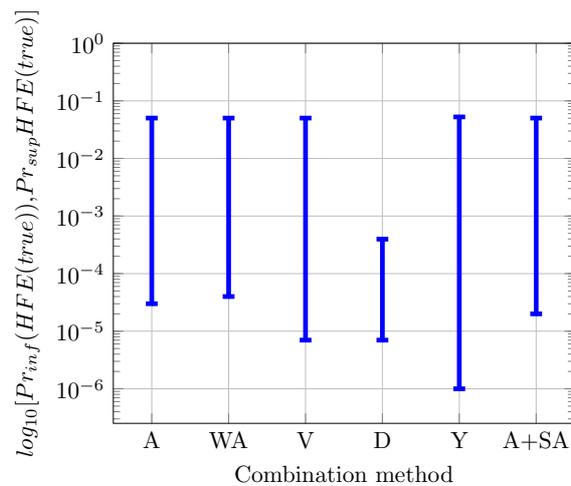


Figure 6.18 – Quantification of the HFE from expert (only) data combination (Average, Weighted Average, Vote, Dempster, Yager’s), and combination of average combined expert data (A) with the average combined simulator data bar plot (A + SA)

combination method used. Here, we select average combination rule (as shown in [Figure 6.17](#)), any combination rule can be selected, and each data combination result generates separate VBS model (similar explanation in [section 5.2.3](#)).

The quantification results are shown in [Figure 6.18](#), the bar-plots show the upper and lower probability bounds, for the HFE being true, i.e.

$$[Pr_{inf}(HFE(true)), Pr_{sup}(HFE(true))].$$

This concludes the usage of objective data in the PRELUDE’s quantitative methodology. The next section details the subjective data’s usage in the PRELUDE methodology.

6.3.3 Subjective data usage: Retrospective identification of PSF self estimation – input to PRELUDE qualitative part

This section analyzes mainly the subjective data to identify PSFs that were either not identified or not considered previously qualitative part of PRELUDE. Such an indicator is essential since not all PSFs are considered in the safety analysis of a human’s objective - top down analysis. Such factors need to be identified from real data. Further to respond to the reason behind inadequate performance - why the performance was inadequate? The component of objective performance in this case the scores also need to be considered. Analyzing the subjective data with the objective data allows us to observe a detailed feedback on the cause of the inadequate performance. This feedback, in this section is interpreted a PSF needs to be considered given the empirical evidence.

The usage of subjective data for the PRELUDE’s qualitative methodology is

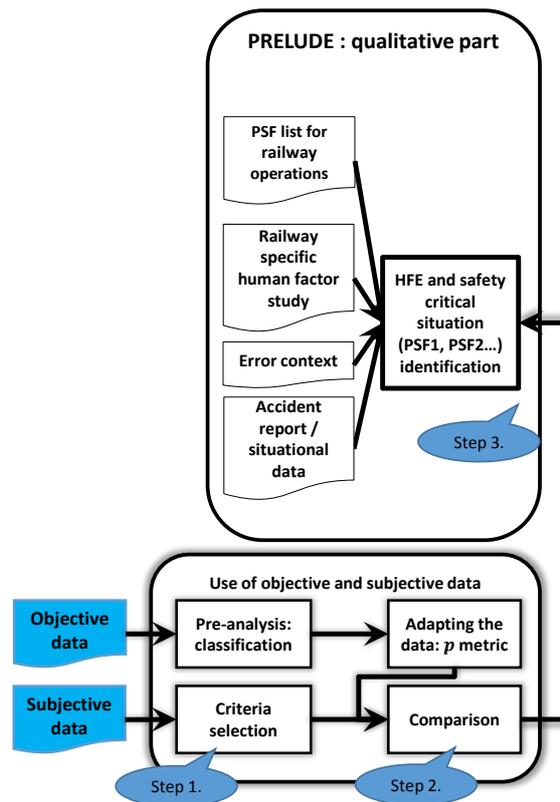


Figure 6.19 – Overview of the usage of subjective data from simulator experimentation in the PRELUDE’s qualitative part

shown in [Figure 6.19](#), an explanation of the two steps follows. In this section, we focus on inter-subject indicators instead of inter-run differences, hence, different parameters for different subjects are compared.

Step 1. Selection of criteria: From the subjective data we select NASA TLX. Mainly because, with the six sub-scales of NASA TLX there is data on varied indicators of a subject’s perception. One of the parameters, we are interested for this section is performance (refer [section 6.2.3.2](#)). Mainly because it represents a dimension - a focus on how a subject evaluates his/her performance. To recall, the question is phrased as: *"How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?"*. Thus, the rating of this parameter, for a subject (averaged over all the runs), is an indication of self perceived notion of success in the given task - in this case a scenario run.

Further, different objective criteria of performance were defined in [section 6.2.3.1](#). These were derived from the principle goals of a train driver, as also explained to the subjects in the explanation sessions in [section 6.2.2.3](#). We consider one of these indicators, the p_{SS} (p metric of the safety score) for a subject averaged over all

the runs. Taking both of these parameters together: the TLX sub-scale rating for performance, and p_{SS} average value for a subject, we have a subject-specific pair. This pair combines the subjective perception of a subject and objective performance measure.

Step 2. Comparison to extract PSF and its states: This pair of values is analyzed here to evaluate *self estimation* of a subject – a PSFs. The two extremes of self estimation, we consider here are *underestimation* or *overestimation*. Suffice to say that both these extremes, are detrimental for accomplishing a task successfully. Thus, these are considered as *poor* states of a PSF - *self estimation*; the other state of this PSF is referred to as *good estimation*. Furthermore, overestimating ones abilities can be also linked to an increased risk taking. A good estimation is linked to support correct performance, that is a *nominal* level of this PSF.

Thus, this pair is used to contrast differences of self estimation. The way these levels are identified for a given subject are described in [Table 6.3](#).

p_{SS} value	TLX performance rating value	Self estimation level
high	high	overestimation
high	low	good estimation
low	high	good estimation
low	low	underestimation

Table 6.3 – Self estimation levels based on Safety score and subjective performance ratings

[Figure 6.20](#) presents on x axis p_{SS} values, and on y axis the performance ratings. As a reminder a high p_{SS} value means a high over-travel distance, and thus the worse safety performance. A data point is the average value of both of these TLX for a subject. Hence, as seen there are 13 data points, once for each subject. Here, as an aid for comparison we add the information on the group of a subject - a purely objective criteria - as identified in [section 6.3.1](#), in the plot.

Some comments using [Table 6.3](#) and [Figure 6.20](#) are made as follows:

- Most of the **group 1 subjects are seen in the underestimation their performance**. Even though their safety score was low, they seem to evaluate their performance below average.
- Group 2 subjects are found in the self evaluating range between Group 1 and Group 3. As can be seen along the y -axis of [Figure 6.20](#): some Group 2 subjects evaluate their performance closer to Group 1 subjects, and some closer to Group 3 subjects. They are aware of their performance. Hence, based

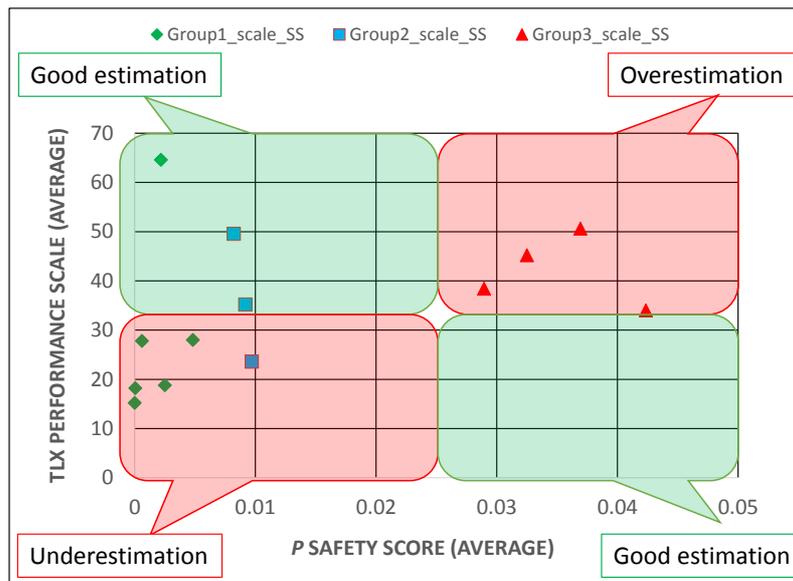


Figure 6.20 – Self estimation level of the subjects using average values of their p_{SS} (the higher the worse for safety), and performance score of TLX, with the three groups of subjects as identified in [section 6.3.1](#)

on this a remark can be that **Group 2 subjects do not necessarily perform the best, but are better at self-estimation vs. Group 1 subjects.**

- **Group 3 subjects are mostly overestimating.** This states that the subjects who are not respecting safety are either - deliberately doing so, or are indifferent of their penalties. In either of the cases, they are overestimating their performance. A possible solution can be in such cases, to re-visit the explanation and or allow subjects to express their interpretation the primary goals, e.g. some subjects might preferred to complete the session on time, and were indifferent to the safety violations they incurred.

Step 3: Integration of the new PSF in the context/qualitative analysis. Contrasting subjective and objective data, as done in this section provides more details into the reasons behind inadequate performance. These reasons are interpreted here as a new PSFs called *self estimation*. This also allows giving possible recommendations to improve performance, by giving concrete perspectives on the possible causes inadequate safe or reliable performance. In the context of PRELUDE this enters the step for the identification of an HFE's safety critical context. It can be concluded that when analyzing over speeding violations subject-wise differences, notably *self estimation* needs to be taken into account. The levels of self estimation degrading to performance were: *overestimation* or *underestimation*, the levels supporting normal performance is *good estimation*. It was also observed that

overestimation generally leads to more safety violations than underestimation. To note that, here, an empirical approach is used for the self estimation level of a subject. A formalization of this process can be proposed to make it easy for application, this can be done for future work.

6.4 Discussion

The protocol: The objective of the experimental protocol was to obtain empirical data, more precisely the effect of a PSF on human performance. In-line with HRA objectives, a focus on the degraded state of the PSFs was kept. However, such a case has some limitations, as stated previously, the considered degraded state of the PSFs were limited to a specific instantiation of the generic definition of the PSF. That is from the generic definition of the PSF, a scenario was defined by modifying some aspects of the scenario run. This was done by either adding additional side-tasks or changing the parameters (additional time constrains, interaction with the environment, etc.), these changes, from a purely qualitative point of view is sufficient for present study's objectives. However, as further robustness measures proof of independences bwtween scenarios can be added to the protocol - by randomizing the order of runs, by using statistical methods of showing independence in teh results - however, such methods might require sufficient data sources and analysis to establish.

The experimental campaign and the protocol: An experimental campaign was carried out with 13 subjects to implement the proposed experimental protocol. It was demonstrated how the data was collected (objective, subjective), and preliminary discussion of the data collected was presented. For a first implementation of such a work a reasonable number of subjects were sufficient. However, more subjects will help in obtaining better data both in terms of quantity and quality. Further, this campaign aimed to demonstrate the feasibility of PRELUDE using simulator data by existing model with more data, and showed more importantly how to obtain such data. Thus, for such activities using university students provides a first trial of such a protocol. More importantly some lessons learned from the campaign are discussed below. Although if one wants VBS models which can be applied to predict human reliability for an industrial usage, as far as experiments on simulator is concerned – real train drivers must be used as subjects.

For the PSF-subjective feedback questionnaire the definitions were at first taken from the PSF list proposed with the PRELUDE methodology. In the case of PRELUDE's case study these definitions were easier for the experts to understand.

However, while responding to the questions the subjects reported some difficulty in understanding these definitions. The definitions were then modified making them simpler and adding more practical examples.

It was also observed NASA TLX requires an effort on the part of the subjects to understand and respond, at least for the first time they respond to the questions. For NASA TLX the subjects sometimes reported difficulty in understanding definitions of some sub-scales and some pairwise comparisons. For some non-obvious ratings, for example physical effort vs. perceived performance. The subjects thus needed additional support for the first time when responding to TLX. This is expected, since previous critiques of TLX have noted the "...lacked internal consistency from the effort and frustration levels reported to the performance scale." [Hankins and Wilson, 1998]. We helped the subjects in these cases by giving them some examples and simpler explanations.

Objective data usage: As it can be observed in [Figure 6.17](#) the values obtained from experts are higher than what is obtained from simulator campaign. This is because the questions asked to the experts were broader in their definitions, mainly the PSFs. Experts were elicited on the complete definitions of the PSFs, whereas simulators runs as proposed in this chapter, is one particular instantiation of a PSF's poor state. We can say that data from experts is a global picture of a PSFs effects on operational safety, whereas, the experimental simulations are able to simulate only a subset of the PSF's definition. For example, a *poor* training/experience in the simulation was limited to knowledge of procedures, knowledge of track, and experience gained by driving of train. Whereas, in reality it can include in addition to these cases, other cases such as: knowledge of train dynamics, experience driving in a particular environment, different knowledge of different procedures etc. An expert takes into account the whole definitions unless otherwise explicitly asked to exclude, even in those case it might be difficult for the expert(s) to visualize such a sub-context. As a possible solution, either experts can be shown data/scenarios of the experimentation and elicited on those particular cases only. Or multiple sub-scenarios that is sub-instances of a PSF poor can be simulated, combined among themselves and then combine with expert data. In such cases we can focus only on certain important PSFs and states of those PSFs.

Subjective data usage: This work also proposed a discussion on identification of PSFs – self-estimation – from experimental data. The subjective and objective data was analyzed to identify it's states for different subjects. This identification was obtained by comparing objective and subjective measures. This new PSF was

not identified as an important PSF in the previously proposed PSF list. It can be in different states, and is a person specific PSF. To note that, here, an empirical approach is used for the self estimation level of a subject. A formalization of this process can be proposed to make it easy for application, this can be done for future work.

6.5 Conclusion

Acquiring data robust and appropriate to model human reliability requires empirical sources, different types, and careful analysis. This work presented one such approach using an operational simulator. It is aimed at capturing the effect of a PSFs' state on human performance by careful preparation of the scenario's conditions. PSFs relevant for the railway domain were considered while designing the simulator session. On one hand, objective criteria which links human performance and the system level goals were chosen, and on the other subjective data was also obtained. Different data sources were used and it was demonstrated how such data can be used in the PRELUDE HRA methodology previously proposed. The simulator set-up of present work is similar to training simulators often used in the railway industry. The use of standard subjective questionnaires, signaling systems such as ERTMS, provide an easily repeatable and usable methodology.

Most PSF lists, such as PSF-lite [Kyriakidis et al., 2015a] take a historical accident analysis point of view to identify and propose what PSF to take into account while analyzing human reliability. However, they do not consider the possibility to obtain PSFs from the usage of simulators which are often used to train operators, as was proposed in the present work. Furthermore, accident analysis on old data can be difficult to apply to current rail operational context. For example, in ETCS context the on-board DMI plays a critical role in the human-machine interaction and a PSF *human machine interface* identified from say 20 year old accident data might not have the same importance. Secondly, such PSF lists are from a singular point of view – safety critical factors from historical data, but what about the dynamic learning effects of PSFs, the identification of new PSF which should be considered due to the changing work environment or task requirements (in-cab driving), or person specific PSFs as identified in present work, etc. Thus, a simulator and subjective data-based approach that this chapter presents allows a more holistic consideration of PSFs in human reliability analysis.

This chapter has thus demonstrated the feasibility of PRELUDE with empirical data from simulator sessions. We have validated that the BFT-VBS framework proposed in PRELUDE is capable of accommodating data from experimental simu-

lations; (2) data from experimental simulation can be used for both quantitative and the qualitative objectives. We have shown that the objective data can support expert data, and subjective data can be used for identification of new PSFs. The simulator data plus expert data-VBS model offers a level of validity greater than only expert or only empirical HRA models. The quantification results obtained from this extended model hence are more representative of reality.

Conclusions and Perspectives

Conclusions

In increasingly complex and evolving transportation system of systems where large resources are allocated towards ensuring operational safety, it becomes necessary to analyze a human and its context, which directly or indirectly influences operations. Systemic accident analysis methods state that accidents are results of emergent behaviors; and a system of systems view states that global safety is an emergent property, thus prevention of accidents and assurance of safety need the use of specific methods. However there is a lack of considerations of human factors alongside other systems in traditional risk and reliability approaches. Further, regulations lack concrete criteria on the risk assessment and acceptance criteria on human factors analysis. Most RAMS analysis although consider human factors and provide guidelines, but these guidelines lack the same level of details as the ones for technical systems.

HRA is a family of methods which can be used to analyze human factors and operational safety. A PSF-based HRA model was identified in this thesis as an appropriate choice. Further, a quantitative HRA was identified as a pragmatic approach, to allow for a unified analysis of human factors and the operational safety of railway operations.

The usage of a complete HRA method (such as PRELUDE) can be more time and resource consuming as compared to purely quantitative methods (such as RARA). However, identification of HFE/PSFs, modeling of the relation between the HFEs and PSFs both qualitatively as an error context and quantitatively from empirical and expert data, and finally quantification as an error probability and feedback on the PSFs – provide results that are more complete and rich in terms of conclusions for improving system safety than other methods.

A critical survey on human error quantification techniques was also performed towards proposing a quantitative HRA methodology for railway application . This discussion was relatively exhaustive in terms of the quantitative frameworks

currently used in the HRA domain. A focus was also made on the data they used to model human reliability and the sources of the said data. Most of the quantitative HRA models proposed in literature use a quantitative model to represent the PSFs influences on the human towards determining an error probability. Some methods use empirical data, but most of them use expert data or data from other models. Furthermore, they agree on the inclusion of a PSF-list or similar set of influencing factors for the analysis. Thus, a rail-specific PSF-list was proposed. We have taken a broader point of view on what PSFs should be included in a HRA analysis.

Our proposition, called the PRELUDE methodology, provides decision-making capabilities to analyze and characterize the probability of occurrence of a human error given an operational context. It can then be applied for a retrospective (to analyze an accident/incident scenario) or prospective (e.g. for a new railway line) analysis approach to ensure the operational safety of rail transportation. The case study using a real-world high speed railway accident where we extracted data from the accident report, was used to demonstrate the usage of PRELUDE for retrospective analysis.

The qualitative part of PRELUDE allows identification of PSFs and HFEs towards an human reliability analysis. These PSF are identified from domain specific human factors and PSF-based studies, and the HFEs come from an analysis of an accident scenario, as the most critical human failure events, which need a detailed analysis. Further, it aims to characterize a safety critical situation as a collection of PSFs which impede a safe accomplishment of a function by a human, significantly more than individual PSF or other sets thereof. This allows for a focus on the context, rather than errors towards a systemic approach to human error analysis.

The quantitative part of PRELUDE models the strength of relations between HFEs and PSFs using configuration belief structure, and the evidence on the individual PSFs using direct belief structures. Configuration belief structure is built based on conditional data: expert or empirical. The PRELUDE chapter details how the experts should be consulted and how to combine their potentially conflicting elicitation into a final valuation. The valuation of the direct belief structures relies on empirical data from accident statistics. The model built as such can be used for further analysis, primarily for determining the human error probability and also for determining which PSFs contributed the most to the HFE. Further, the PRELUDE quantitative framework is flexible: that is the variables HFE and PSF and their states are not a limitation to the framework. Both in terms of the number of states and the way in which the relations between them is defined, this makes it a flexible approach. Thus, it can adapt based on the nature of the data available, which might be different based on its source (expert, experimental, accident statistics, etc.), and

formulate it as a configuration belief structure and integrate into the VBS model currently proposed.

Most simulator experimental campaigns for HRA objectives focus on obtaining the frequency of human error data. This work proposed a PSF-centered protocol which aims to simulate particular states of PSFs. This protocol aims to capture the influence of the PSF's states on human performance. Thus, the focus on context – states of PSFs – to obtain empirical data makes the proposed approach more robust and accurate than classical human error frequency-based approaches. Further, the proposed experimental protocol contributes towards addressing the lack of data in human reliability problem from a purely data collection standpoint. The simulator set-up of present work is similar to training simulators often used in the railway industry. The use of standard subjective questionnaires and signaling context such as ERTMS, provide an easily replicable and usable methodology.

As was demonstrated, it is feasible to use the PRELUDE methodology and combine both expert and simulator data to build the model. Note that these two activities are completely independent of each other. PRELUDE methodology's feasibility to take into account data from such different sources is another positive point of our proposition of the BFT-VBS framework.

Validation is indeed an inherently problematic issue for most HRA models. Most models are largely based on expert opinion and there are no HRA benchmarks that can be used to validate these models. Thus, we have demonstrated the feasibility of PRELUDE by augmenting it with data from an experimental simulation campaign. Using both expert and simulator data in the same model provides a level of validity greater than only expert or only empirical HRA models.

Perspectives

PRELUDE methodology in its current state is developed for, and applied to a case study for the railway domain. But it may very well be, with moderate effort applied and adapted to other transportation SoS. A completely genetic HRA model is not only rare (see [Table 4.3](#)), but is often inspired from notions of previous domain-specific HRA model.

The present quantitative approach of PRELUDE models takes as input the data on $PSF - HFE$ relation, that is the effect of context on a human. This data is obtained from the simple questions to experts, and simulator sessions, this data is then combined as a second step to represents the combined effect of a context with multiple PSFs, giving the relation $PSF(s) - HFE$. Explicit input data concerning the relation between multiple PSFs, is not sought in the current work.

Following rather evidently, questions be asked from the experts, or experimental simulations can be carried out where a combined influence of several PSFs is sought. Formulating such questionnaires, and carrying out such simulator sessions and ensuring that appropriate data is obtained needs more work. However, such data can then be integrated in the formal transformation approach currently proposed in [section 5.1.2.3](#) PRELUDE to build VBS models.

Further, in the current version of the configuration belief structure pure *PSF* – *PSF* dependencies are not considered. A PSFs influence on another PSF can be modeled using an intermediate belief structure to account for common cause effects. For example a belief structure which defines the relation between Task Load and Communication; which then links to the central configuration belief structure. Such a decomposition can be proposed to represent inter-PSF relations. The proposed VBS-BFT framework can express such relations between the variables.

Expert data remains the primary choice for an HRA methodology, since it is relatively easier to obtain, as compared to extensive experimental campaigns. At first level, careful work needs to be done to prepare the elicitation process. In a preliminary discussion, feedback or remarks of the experts or analysts on the structure of questionnaire can be obtained to improve the elicitation process (questions, context details, responses: probability or qualitative responses, importance of factors, etc.), mainly to ensure that the analyst and experts have the same understanding towards assuring the accuracy of obtained data. The questions can be formulated so that they allow taking in account the confidence an expert in the responses given, thereby giving the analyst means to accurately represent expert(s) knowledge in the model. Secondly, the need of formal ways to combine expert information can help in increasing the repeatability and transparency of the process. However, it remains an important issue, and it can be argued that it is a philosophical rather than a mathematical problem. The hypothesis presented by a combination rule and the subsequent choice thereof is not straightforward, to support this PRELUDE presents an open discussion of the different combination methods. Some results on the combination of conflicting expert opinions and how to manage the conflict.

Further, empirical data from simulators can also be used to support the elicitation process. An iterative approach can be adopted which uses simulator data as support for the elicitation process. It is then seen as a complementary approach for collection data on the PSFs. The experts have access to empirical data to allow them to make inferences from this data, to support their arguments or to make them reconsider their opinions. Along similar lines, there is a complementary nature of expert and simulator data, this needs to be exploited further. For example,

it is relatively easy to create a simulator run with high task load and observe human performance. On the other hand expert opinions on this influence can vary, whereas they can provide an account of how a good experience influences human performance.

More experimental campaigns can be carried out with the proposed simulator protocol to obtain more data. As a rather straightforward approach only objective data from the experimental protocol presented in this work can be used to build a quantitative HRA model. A focus can also be made beyond the current considerations of degraded levels: e.g. how the performance is influenced when there is a good level of communication, that is to consider the positive effects of PSFs.

The data that was collected in the experimental campaign can be further exploited towards other PSF-based analysis. For example, we need to analyze the dynamic and static nature of a PSF, i.e. initial experience and experience after a learning phase that is after having run the simulation. These behaviors should be added to the differences in performance between the runs coming from degraded PSFs states. To mitigate this effect, the run order can be randomized, i.e. different subjects could be asked to do the runs in different order.

System level, multi-criteria risk-based inferences, and taking account of positive aspects of PSFs and human actions can be appended to PRELUDE towards an extended methodology. This is aimed at further easing decision-making capabilities, and performing a holistic analysis of a human error.

Appendix

A.1 Appendix to the PRELUDE methodology

A.1.1 The flowchart representation of the PRELUDE methodology

This flowchart follows the explanation to the PRELUDE methodology as described in the [section 5.1](#). It aims to present a more clear user-oriented illustration of the PRELUDE methodology. The rest of this page is left blank intentionally.

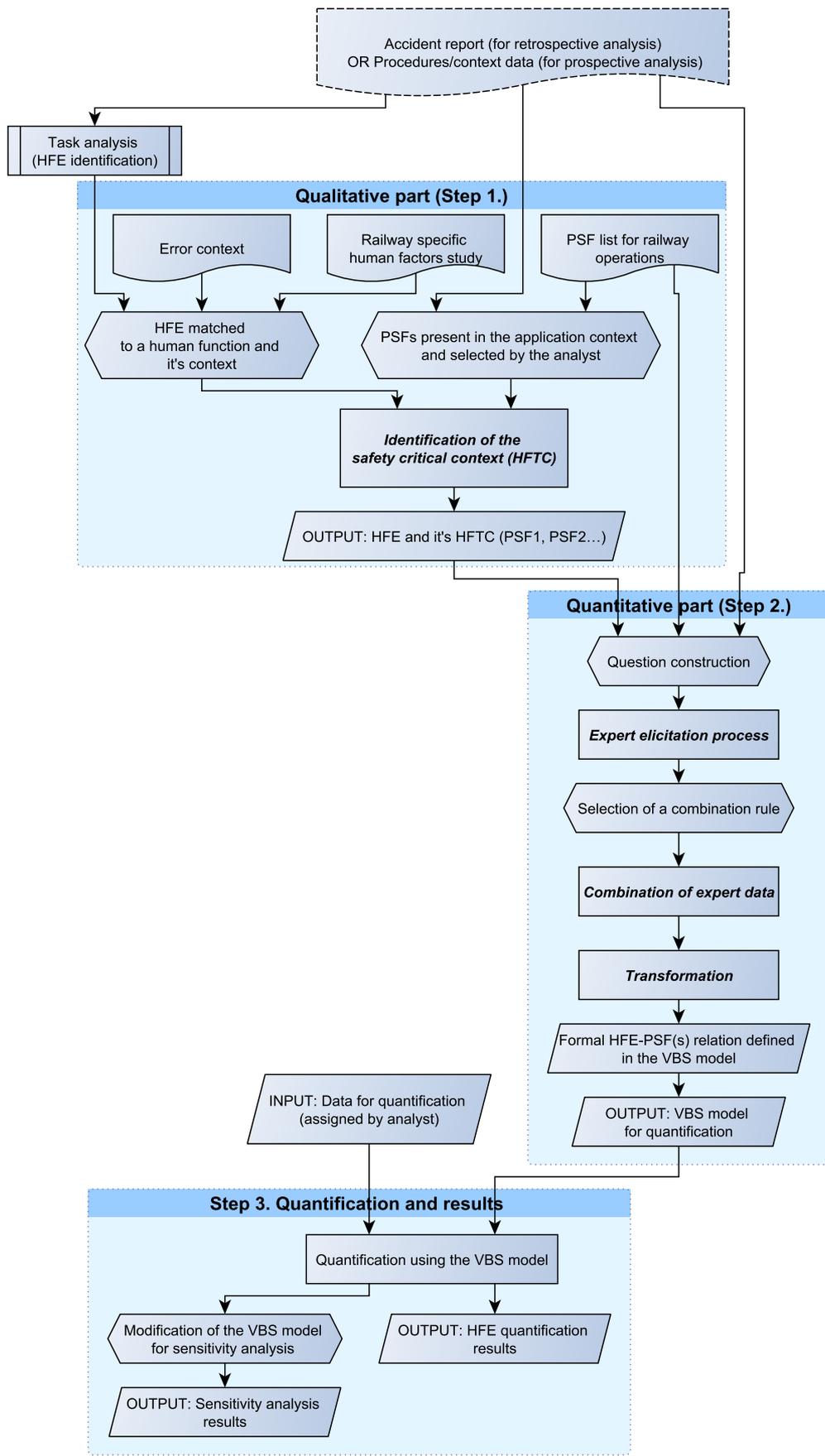


Figure A.1 – A flowchart description of the PRELUDE methodology.

A.1.2 Case study: extraction of data from the accident report

The accident investigation report was analyzed in order to extract the PSFs and the HFEs. The raw data that was extracted from the accident report (later refined for the case study) is provided in the table below.

Location ¹	Event	Cause	Procedures / Contextual information
X	Excessive speed - 180km/h	Not respecting speed signals / failure to brake to slow train to acceptable speed.	Any agent, regardless of its function, is to obey passive and immediate the signals concerning him.
x-90m x-10sec	Phone call longer than normal duration (Ineffective Communication)	- "The conversation has an excessive duration (100 seconds)..." - 5540 meters distance traveled since the start of conversation "to inquire about the stop at the next station".	Absence of firm guidelines (Guide to good practice in Driving neither regulatory nor policy).
x-200m x-13sec	Not respecting the Speed Signals (speed marked in in schedule book)	- "...lack of attention by the driving personnel..." - Schedule book - "speed change at mark 84.2 from 220 km/h to 80km/h.	maximum speed change marker boards (Image below) indicates the point on the line where the maximum speed allowable by infrastructure changes, as established in the tables of maximum speeds. No regulations as to what type of signaling preventive fixes should be implemented in maximum speed changes. The system helps driving ASFA not provide oversight, and therefore drivers must follow the directions shown by the lateral signaling, risk exported to the driver. 
x-230m x-14sec			Speed indication in the Audible sound from the ASFA balise only if the track ahead is not free (balise for signal E7 see image below). 

¹ (x is the point of accident) in distance (m) and time (sec)

x-4000m	Not braking at usual braking point	- reference point for breaking (signal E'7, see image) not observed due to a lack of attention by the driving personnel" - Inadequate perception of environment to identify breaking point.	
x-4100m			Approaching signal with reduced visibility 
x-4300m x-68sec			Audible sound from the ASFA balise only if the track ahead is not free (balise for signal E7 see image below). Dead man's switch actioned.
x-4300m			Exist from ERTMS zone (ETCS nonfunctional). (nonfunctional) Transition to STM (Specific Transition Module) from CAB signaling ERTMS is accomplished by a sound signal in the cabin indicating the driver proximity to a transition point; Acknowledgement from driver required in a duration of 5 seconds, automatic brake application in case of non-acknowledgement. 
x-6000m x-110sec		Start of phone call.	
...			
x-23 to x-10 minutes	Recall later - "Do not know what I was thinking before entering the tunnel."		
x-25 minutes	Automatic braking by 'dead man's switch' (two times prior to the accident.	Due to failed to press DMS OR fault of system. Although the driver makes use of the pedal immediately after the acoustic signals, occurs anyway emergency braking.	

Other information

Experience	The train driver "made the same journey several times a week"; 2 years of high speed train driving.
Training	"Driver complied with current regulations regarding title, revision of training, medical examinations and qualifications (rolling stock and line)." The danger posed by the curve of Angrois (accident point) was treated exclusively in the field training, but not shared with safety management.
Signaling	ASFA, line-side signal and driver's book. (ERTMS switched off due to maintenance. Switched off one week after / possible cause 1. different versions of ERTMS which led to failure to read balise, and automatic change of ETCS to mode SR instead of FS).
State of driver	Total driving time of time 2 hours and 44 minutes, the work day of the driver was 8 hours and 47 minutes. "All periods of time (working hours, driving and rest) are made by the driver within current regulations." Test of alcohol, drugs and medicines yielded negative results. "Working hours and driving times and rest - meet with current regulations."
	Inspections of cab (ob-board systems) on line 082 from its commissioning not have detected abnormalities or incidents in the infrastructure, no anomaly in the section of track in the area of the accident.

SOURCES

Comisión de investigación de accidentes ferroviarios. (2014). Informe final sobre el accidente grave ferroviario no 0054/2013 ocurrido el día 24.07.2013 en las proximidades de la estación de Santiago de Compostela (A Coruña). Retrieved from <http://www.fomento.gob.es/NR/rdonlyres/0ADE7F17-84BB-4CBD-9451-C750EDE06170/125127/IF240713200514CIAF.pdf>

ASFA http://www.camins.cat/emails/Cursos/Curs_ferrovieri_2014/ponencies_web/25.4.14/Senyalizacio_proteccion_Cantero_25.4.14.pdf
; http://www.sindicatoferrovieri.com/DOCUMENTACION/CIRCULACION/RGC_2006b.pdf

RFF (Réseau Ferré de France) now SNCF Réseau. (n.d.). *Principes et règles d'exploitation du système ETCS - Particularités en cas de superposition à un autre système de signalisation Document d ' exploitation - RFN-IG-SE 02 C-00-n°002- Version 01 du 09-12-2013.*

A.2 Experimental campaign questionnaires, source code and additional data collected

A.2.1 NASA Task Load Index (NASA-TLX) Questionnaire

The first type of questionnaire is used for a subjective estimation of workload - NASA TLX [Hart and Staveland, 1988]. NASA TLX stands for National Aeronautic and Space Administration task load index. The authors describe workload in ". . . workload is not an inherent property, but rather it emerges from the interaction between the requirements of a task, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the operator." It is described as a workload measurement technique presenting empirical validation supporting it.

Generally, it is known to be a good multidimensional scale for measuring mental workload. It was stated to be "consistently superior" in terms of sensitivity to changes as measured by factor validity (correlation with the workload factor), and operator acceptance [Hill et al., 1992]. It has been used across in various domains and over the years [Hart, Sandra, 2006]. Including some HRA-related objectives [Ha and Seong, 2009].

This method is based on six semantic descriptors (or dimensions or sub-scales) of workload. Namely: *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*. We will use sub-scales henceforth, to refer to these descriptors. The sub-scales are defined in [Hart and Staveland, 1988], and are also presented to the subjects while responding. The definitions of these sub-scales are given in Table A.1, followed by a brief explanation of the subsequent calculation to compute overall workload rating.

NASA TLX for each dimension asks for a rating on bipolar scale. These values are refereed to as sub-scale score or *scales*. The second step, is a weighting process that requires a paired comparison task. The subject to choose which dimension is more relevant to workload for a particular task across all pairs (15) of the six dimensions. Finally, the workload scale (or TLX rating) is obtained for each task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by the total weight. It is given by the following equation:

$$TLX \text{ Score} = \frac{1}{15} \sum_{i=1}^6 D_i \times C_i \quad (\text{A.1})$$

where: *TLX score* is the global workload score, D_i is the raw sub-scale rating for the descriptor i , and C_i is the number of times a descriptor was chosen in the

Sub-scale descriptor	Definition	Rating scale score
Mental demand	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?	1 to 100
Physical demand	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?	1 to 100
Temporal demand	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?	1 to 100
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?	1 to 100
Frustration	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	1 to 100
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	1 to 100

Table A.1 – NASA TLX descriptors: questions and scale

pairwise comparisons. The sub-scale scale score or the *TLX Scores* can be used when evaluating or comparing a task. The output are TLX's individual sub-scale values, respective weights (after the pairwise comparison), and finally the TLX rating (or TLX global score). These outputs will be used in the analysis of subjective data, as described later in this chapter.

A simple feedback used to is often used to HRA data collection studies from simulators (e.g. operator PSF ratings in [Skjerve and Bye, 2011]). From the most evident advantage of being simple, quick and straightforward to take, and to post process.

Along similar lines also towards HRA objectives, in [Hallbert et al., 2012], the authors propose self-rating on 5 point rating scale (on the PSFs), with 1 being a hindrance to the performance and 5 helping the perform better. This aimed to ask the operators to rate the PSFs "... in terms of their perceived influence on their performance." Similar 5 point scales were used in [Bareith and Karsa, 2009].

A literature review and other such methods (Modified Cooper-Harper Scale, Subjective Workload Assessment Technique, etc.) can be found in the literature

review of workload measures [[Miller, 2001](#)].

A.2.2 Subjective questionnaires: Pre, post and PSF subjective questionnaire

This section gives the questionnaires that were used to obtain subjective data, that is the PSF-subjective questionnaire and the pre and post questionnaire are given. The rest of this page is left blank intentionally.

PRE Questionnaire SUBJ-1.

To fill before starting the session. Select one of three options...

*Required

1. **Participant ID:** *

2. **PRE1. Did you understand the explanations given to you?** *

Mark only one oval.

- yes, completely
 no, very few things
 Most of the things

3. **PRE2. Do you understand your main objectives of this experimentation ?** *

Mark only one oval.

- yes
 No
 Most of it, yes

4. **PRE3. In general, are you able to understand a new situation quickly and be aware of what is happening ?** *

Mark only one oval.

- Yes, most of the times
 No, it takes me time to adjust
 I am not sure

5. **PRE4. Are you able to do multiple tasks (multitask) in a given time?** *

Mark only one oval.

- yes, most of the times
 no, it takes me time to adjust
 I am not sure

6. **PRETLX1. Are you physically tired?** *

Mark only one oval.

- yes
 No
 I am not sure

7. PRETLX2. Are you mentally tired? *

Mark only one oval.

- yes, I am tired
- no, I am completely attentive
- I am not sure

8. PRETLX3. Do you think you can perform all the scenarios that will be given to you successfully (based on the explanation...)? *

Both presented and pre-read material

Mark only one oval.

- Yes
- No
- Maybe

9. PRETLX4. Are you ready to put all your effort in the task? *

Mark only one oval.

- Yes
- Maybe
- I am not sure

10. PRETLX5. Do you easily get frustrated? *

Mark only one oval.

- Yes, if I cannot do what I am supposed to do
- no, I will try next time
- depends

Questionnaire SUBJ-2

Select one of the three options.

*Required

1. **Participant ID: (ask admin) ***

2. **RUN number: (Ask the admin) ***

3. **SUBJ.2.1 - In your opinion were the conditions/situations in the simulation real?**

Mark only one oval.

- yes
 no
 I am not sure

4. **SUBJ.2.2. In your opinion how difficult was the scenario? (compared to others)?**

Mark only one oval.

- easy
 normal
 hard

Questionnaire SUBJ-2.

Rate the following factors (PSFs) based on if they helped, did not affect, made worse your performance in the session just performed.

Training

Did you had all the correct knowledge (from the explanation session) to do what you were asked to do?

All the signals, procedures, signal boards, etc.

5. **Training ***

Mark only one oval.

- I had a good training - it improved my performance.
 I had a nominal training - it made me perform correctly, but did not improve it.
 I had a poor training - it made my performance worse
 I am not sure

Experience

Information and knowledge that you have by doing the same thing.
You know what to expect and what to do, because you have seen it before.

6. Experience *

Mark only one oval.

- I had a good Experience - it improved my performance.
- I had a nominal Experience - it made me perform correctly, but did not improve it.
- I had a poor Experience - it made my performance worse
- I am not sure

Communication

Did you communicate (with Agent , or controller, if needed) well, or some communication that you received helped you in doing something. If you did not communicate with anyone, chose second option (nominal).

7. Communication *

Mark only one oval.

- I had a good Communication - it improved my performance.
- I had a nominal Communication - it made me perform correctly, but did not improve it.
- I had a poor Communication - it made my performance worse
- I am not sure

Situational Awareness

You know what was happening in the scenario run.
You were also able to predict what was going to happen.
You were completely concentrated/attentive in the task

8. Situational Awareness *

Mark only one oval.

- I had a good Situational Awareness - it improved my performance.
- I had a nominal Situational Awareness - it made me perform correctly, but did not improve it.
- I had a poor Situational Awareness - it made my performance worse
- I am not sure

Task Load (Workload)

The tasks that were assigned to you were not too much in number.
They were also not too complex for you to perform.
NOTE: This includes only the main tasks of driving, e.g. playing game does not count as task load.

9. Task Load (Workload) **Mark only one oval.*

- I had a good Task Load (Workload) - it improved my performance.
- I had a nominal Task Load (Workload) - it made me perform correctly, but did not improve it.
- I had a poor Task Load (Workload) - it made my performance worse
- I am not sure

Time load (Workload)

Did you had enough time to complete tasks that you were supposed to do?
Did you had too many tasks to do in too less time?

10. Time load (Workload) **Mark only one oval.*

- I had a good Time load (Workload) - it improved my performance.
- I had a nominal Time load (Workload) - it made me perform correctly, but did not improve it.
- I had a poor Time load (Workload) - it made my performance worse
- I am not sure

Human system interface (HSI) quality

Was the quality of the DMI good?
Did it display all the relevant information?
Was I able to see (visibility) and hear (audio/sounds) things relevant to what I was doing?

11. Human system interface (HSI) quality **Mark only one oval.*

- The HSI was good - it improved my performance.
- The HSI was nominal - it made me perform correctly, but did not improve it.
- The HSI was poor - it made my performance worse
- I am not sure

POST Questionnaire SUBJ-3.

To fill after completing all the session runs. Respond based on how you think you performed in all the simulation runs.

Select one of three options...

*Required

1. Participant ID: *

.....

2. POST1. Did you understand the explanations given to you? *

Mark only one oval.

- yes, completely
 no, very few things
 Most of the things

3. POST2. Do you think you understood main objectives, what you should do ? *

Mark only one oval.

- yes
 No
 Most of it, yes

4. POST3. Were you able to understand new situations (signals, tasks, etc.) and be aware of what is happening? *

Mark only one oval.

- Yes, most of the times
 No, it took me time to adjust
 I am not sure

5. POST4. Were you able to multitask in the given scenario? *

Mark only one oval.

- yes, most of the times
 No, few times only
 I am not sure

6. POSTTLX1. Are you physically tired? *

Mark only one oval.

- yes
 No
 I am not sure

7. POSTTLX2. Are you mentally tired? *

Mark only one oval.

- Yes
 No
 I am not sure

8. POSTTLX3. Do you think you performed most of the tasks in scenarios successfully? *

Both presented and pre-read material

Mark only one oval.

- Yes
 No
 Maybe

9. POSTTLX4. Did put all your effort in the tasks? *

Mark only one oval.

- yes, fully in most of the tasks
 no, I might have tried harder
 I am not sure

10. POSTTLX5. Did you get frustrated? *

Mark only one oval.

- yes, for most of the scenario runs
 no, very few times
 equally frustrated and not frustrated

A.2.3 Code used to extract data from Eurocab.log

The *Eurocab.log* file contains, for a complete scenario data about some essential parameters. An excerpt from this log file is given as follows:

```
>(141.93 s, 1089.38 m) – SPEED – front end location = 1087 m, train speed = 97 km/h (P =
    160 km/h, W = 169 km/h, SBI = 173 km/h, EBI = 191 km/h)
>(141.93 s, 1089.38 m) – TARGET – 2344 m, 80 km/h, Pre-IP = 229 m
>(141.93 s, 1089.38 m) – EOA – 10392m, 0km/h
>(141.94 s, 1089.38 m) – RADIO – TX, RBC = 1234, PhoneNb = 12345678FFFFFFFF DATA
    NID_MESSAGE = 132, content = 84 06 80 00 0D DA 40 00 00 40 00 81 40 00 0C 41 5E 00
    01 E0 03 C8 0C B2 60 30
>(142.54 s, 1104.47 m) – SPEED – front end location = 1102 m, train speed = 97 km/h >(P =
    159 km/h, W = 168 km/h, SBI = 173 km/h, EBI = 190 km/h)
>(142.54 s, 1104.47 m) – TARGET – 2344 m, 80 km/h, Pre-IP = 229 m
>(142.54 s, 1104.47 m) – EOA – 10392m, 0km/h
```

For a given data point " $(t \text{ s}, d \text{ m})$ - SPEED - front end location = $fel \text{ m}$, train speed = $ts \text{ km/h}$ ($P = p \text{ km/h}$, $W = w \text{ km/h}$, $SBI = sbi \text{ km/h}$, $EBI = ebi \text{ km/h}$)." These TLX and more are described below, in order:

- t time: a timestamp (in seconds) which counts incrementally from the start of the scenario
- d distance: in meters, the distance traveled by the train from the start of the scenario and 'front end location' an ETCS parameter used to locate the head of the train.
- fel front end location: the location of front end of train
- speed: in km/h , train speed ts corresponding to all of the ETCS braking curves (more explanation in the next section)
- driver interaction (acknowledgment, data entry etc.)
- on-board and signaling data (automatic brakes, RBC messages, EOA location)

Thus, basic parameters of train position and speed can be extracted from this file. It can also be used to extrapolate other parameters such as acceleration, driver reaction time, etc. To ease data collection a bash script was written, which asks for the subject and other identifying information and automatically saves this log file for each run. This log file is the raw objective data, saved specific to each simulation that will be run.

This raw file is then fed to the following python script which then extracts the relevant data and calculates the safety scores as detailed in [REFER chapter 4]

Listing A.1 – Python script to extract and generate safety score for each subject's run

```

#!/usr/bin/python

"""
Author: Subeer RANGRA
DATE:30/04/2017
Version: v3Clean
TITLE: Safety Score from Eurocab.log
"""
def speed_curve_score(inputfilename: object, outputfilename: object, runnumber: object,
                      subjectnumber: object) -> object:
    """
    INPUT: Eurocab.log file location: inputfilename
    OUTPUT: Eurocab.csv for each SUBJECT : with time, distance, raw speed curve score and other parameters.
    WHAT: This takes the raw eurocab.log file and extracts the following, for a subjects each run i.e. Subject/Run/Eurocab.log
    Safety Score: integral_final_speed_curve_score,
    brakedown of the speed curve score :
        'Permissible Speed Curve sc':excessSpeed_permissible,
        'warning speed ':excessSpeed_warning,
        'SBI speed ':score_sbi,
        'EBI speed ':excessSpeed_ebi
    into one single csv file, for each subject for manual/EXCEL.

    ARGUMENTS:
        inputfilename: Eurocab.log location for a subject's run,
        outputfilename: Single output CSV file name as output filename,
        runnumber : Run number for a subject number as run number
        subjectnumber : subject number as subject number
    """

    inputfile = open(inputfilename)
    outputfile = open(outputfilename, 'a')
    if subjectnumber == 0:
        if runnumber == 1:
            # Output only the score parameters

            outputfile.write(
                "Subject_Number" + ";" + "Run_Number" + ";" + "IntegralF_SCS_-_Wnormalized--SPEEDParameters" + "\n")

            # complete set of data output
            """
            outputfile.write(
                "Subject Number" + ";" + "Run Number" + ";" + "IntegralF_SCS - Wnormalized - SPEEDParameters" +
                ";" + "Global Final Time Score" + ";" + "Global Discreet Final Time Score" +
                ";" + "Global speed_curve_score - count" + ";" + "final_total_distance " +
                ";" + "sum delta t for permitted" + "\n")
            """

        index = -1
        test_counter = 0
        count_excessSpeed_permissible = 0
        count_excessSpeed_warning = 0
        count_excessSpeed_sbi = 0
        count_excessSpeed_ebi = 0

        # SET FLAGS for the count
        flag_permissible_speed_count = 1
        flag_warning_speed_count = 1
        flag_sbi_speed_count = 1
        flag_ebi_speed_count = 1

        final_area_v_permissible = 0
        final_area_v_warning = 0
        final_area_v_sbi = 0
        final_area_v_ebi = 0

        sum_delta_t_permissible = 0

        dVWarning_min = 1
        dVSBI_min = 1
        dVEBI_min = 1

        dVWarning_max = 0
        dVSBI_max = 0
        dVEBI_max = 0

        final_time = 0

```

```

# Normalized Weights From distance Parameters --
weight_permmissible = 0.1
weight_warning = 0.2
weight_SBI = 0.3
weight_EBI = 0.4

#speed conversion to meters per second
convert_kmh_ms = 0.2777
#for deltaT
temp_time_iminus1 = 0
temp_time = 0
# for global score g_
g_count_speed_curve_score = 0
integral_final_speed_curve_score = 0
final_total_distance = 0

# score is calculated and summed up for each line -- speed data in the Eurocab.log file
# i.e. each instance of speed measures for a DELTA t
for line in inputfile:
    # i=i+1
    index = line.find("SPEED")
    if index > 0:
        # finding TIME
        index_distance = line.find("")
        index_length = len("")
        index_distance_end = line.find("s", index_distance)
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing TIME
        temp_time_iminus1 = float(temp_time)
        temp_time = float(temp_write_string)

        # finding DISTANCE
        index_distance = line.find("front_end_location_=")
        index_length = len("front_end_location_=")
        index_distance_end = line.find("m", index_distance)
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing DISTANCE
        temp_distance = int(temp_write_string)

        # finding TRAIN SPEED
        index_distance = line.find("train_speed_=")
        index_length = len("train_speed_=")
        index_distance_end = line.find("km/h", index_distance)
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing Train speed
        temp_train_speed = int(temp_write_string)

        # finding PERMISSIBLE SPEED
        index_distance = line.find("P_=")
        index_distance_end = line.find("km/h", index_distance)
        index_length = len("P_=")
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing permissible speed
        temp_permmissible_speed = int(temp_write_string)

        # finding WARNING SPEED
        index_distance = line.find("W_=")
        index_distance_end = line.find("km/h", index_distance)
        index_length = len("W_=")
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing warning speed
        temp_warning_speed = int(temp_write_string)

        # finding SBI -- service brake speed
        index_distance = line.find("SBI_=")
        index_length = len("SBI_=")
        index_distance_end = line.find("km/h", index_distance)
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing SBI speed
        temp_sbi_speed = int(temp_write_string)

        # finding EBI -- emergency brake speed
        index_distance = line.find("EBI_=")
        index_length = len("EBI_=")
        index_distance_end = line.find("km/h", index_distance)
        temp_write_string = line[index_distance + index_length:index_distance_end]
        # storing EBI speed

```

```

temp_ebi_speed = int(temp_write_string)

#TEMP -- to remove
dVWarning = temp_warning_speed - temp_permissible_speed
dVSBI = temp_sbi_speed - temp_permissible_speed
dVEBI = temp_ebi_speed - temp_permissible_speed

if dVWarning >= dVWarning_max:
    dVWarning_max = dVWarning
if dVWarning < dVWarning_min and dVWarning != 0:
    dVWarning_min = dVWarning

if dVSBI >= dVSBI_max:
    dVSBI_max = dVSBI
if dVSBI < dVSBI_min and dVSBI != 0:
    dVSBI_min = dVSBI

if dVEBI >= dVEBI_max:
    dVEBI_max = dVEBI
if dVEBI < dVEBI_min and dVEBI != 0:
    dVEBI_min = dVEBI

# COMPUTING SPEED SCORE FOR EACH PARAMETER...
if temp_train_speed - temp_permissible_speed >= 0:
    # COUNT SCORE
    if temp_train_speed == temp_permissible_speed and flag_permissible_speed_count == 1 and temp_train_speed > 0:
        count_excessSpeed_permissible += 1
        # RESET FLAG
        flag_permissible_speed_count = 0
    # INTEGRAL SCORE

    delta_v_permissible = temp_train_speed - temp_permissible_speed
    delta_t = temp_time - temp_time_iminus1
    #HERE MESURING THE AMOUNT OF TIME IT PASSED PERMITTED SPEED -- FOR 08/02/2017
    sum_delta_t_permissible += delta_t
    area_v_permissible = delta_v_permissible * delta_t
    final_area_v_permissible += area_v_permissible

if temp_train_speed - temp_warning_speed >= 0:
    # COUNT SCORE warning
    if temp_train_speed == temp_warning_speed and flag_warning_speed_count == 1 and temp_train_speed > 0:
        count_excessSpeed_warning += 1
        # RESET FLAG
        flag_warning_speed_count = 0

    # INTEGRAL SCORE
    delta_v_warning = temp_train_speed - temp_warning_speed
    delta_t = temp_time - temp_time_iminus1
    area_v_warning = delta_v_warning * delta_t
    final_area_v_warning += area_v_warning

if temp_train_speed - temp_sbi_speed >= 0:
    # COUNT SCORE
    if temp_train_speed == temp_sbi_speed and flag_sbi_speed_count == 1 and temp_train_speed > 0:
        count_excessSpeed_sbi += 1
        # RESET FLAG
        flag_sbi_speed_count = 0
    # INTEGRAL SCORE

    # FOR INTEGRAL SCORE
    #score is added incrementally -- for all the curves
    #difference -- train speed and EBI speed -- delta_v_ebi
    delta_v_sbi = temp_train_speed - temp_sbi_speed
    delta_t = temp_time - temp_time_iminus1
    area_v_sbi = delta_v_sbi * delta_t
    final_area_v_sbi += area_v_sbi

# if train speed greater than equal to EBI
if temp_train_speed - temp_ebi_speed >= 0:

    # FOR COUNT SCORE
    # not not count only next time -- when go below and go back up...
    if temp_train_speed == temp_ebi_speed and flag_ebi_speed_count == 1 and temp_train_speed > 0:
        count_excessSpeed_ebi += 1
        flag_ebi_speed_count = 0

    # FOR INTEGRAL SCORE

```

```

#score is added incrementally -- for all the curves
#difference -- train speed and EBI speed -- delta_v_ebi
delta_v_ebi = temp_train_speed - temp_ebi_speed
#delta t
delta_t = temp_time - temp_time_iminus1
#area under the curve -- deltaVEBI*deltaT
area_v_ebi = delta_v_ebi * delta_t
#storing in a variable which holds the final deltaVEBI
final_area_v_ebi += area_v_ebi

# SET ALL FLAGS
# if the train speed goes back to being less than EBI/SBI/WARNING/PERMITTED speed
if temp_train_speed < temp_ebi_speed:
    # set flag back to 1
    flag_ebi_speed_count = 1
if temp_train_speed < temp_sbi_speed:
    # set flag back to 1
    flag_sbi_speed_count = 1
if temp_train_speed < temp_warning_speed:
    flag_warning_speed_count = 1
if temp_train_speed < temp_permissible_speed:
    flag_permissible_speed_count = 1

# Since time and distance are stored as incremental values in the Eurocab.log file
if temp_train_speed > 0:
    final_time = temp_time
if temp_train_speed > 0:
    final_total_distance = temp_distance

# here all lines have been parsed -- all the score parameters are combined
# MULTIPLIERS HERE instead of in the excel file

integral_final_speed_curve_score = weight_permissible * final_area_v_permissible + \
    weight_warning * final_area_v_warning + weight_SBI * final_area_v_sbi + \
    weight_EBI * final_area_v_ebi

FINAL_TIME_AVERAGE = 720
final_time = float(final_time)
if final_time <= FINAL_TIME_AVERAGE:
    g_final_time_score = 0
    # Discreet -- means if arrived on time -- success of mission = 1 ; if late failure of mission = 0
    g_discreet_final_time_score = 0
else:
    g_discreet_final_time_score = 1
    g_final_time_score = (final_time - FINAL_TIME_AVERAGE) / FINAL_TIME_AVERAGE * 10000

inputfile.close()

#writing in the CVS file
#print('OUTPUT g_count_speed_curve_score', g_count_speed_curve_score)
outputfile.write(str(subjectnumber) + ";" + str(runnumber) + ";" + str(integral_final_speed_curve_score) + "\n")

# to output all the parameters
"""
outputfile.write(str(subjectnumber) + ";" + str(runnumber) + ";" + str(integral_final_speed_curve_score) +
    ";" + str(g_final_time_score) + ";" + str(g_discreet_final_time_score) +
    ";" + str(g_count_speed_curve_score) + ";" + str(final_total_distance) +
    ";" + str(sum_delta_t_permissible) + "\n")
"""

print('sum_of_all_of_the_delta_V_for_permitted_--_sum_delta_t_permissible', sum_delta_t_permissible)
outputfile.close()

#clean temp csv files
def remove_temp_csv_files(dir_to_remove_files):
    import os
    for file in os.scandir(dir_to_remove_files):
        if file.name.endswith(".csv"):
            os.unlink(file.path)

def main():
    import os
    NumberOfSubjects = 13
    print('the_number_of_subjects_are_{}_d' % NumberOfSubjects)
    output_path = 'D:/GoogleDrive/1_OFC_Work/work/1-2_Thesis/code_and_data/Chapter4_exp/DATA/OPSIMU/analysis/'
    # Detecting and if exists -- deleting temp (old files)
    if os.listdir(output_path) != []:
        print('there_are_temporary_files,_do_you_want_to_delete_??')

```

```

print(output_path)
keystroke = input('press_enter_to_delete_files,_a_value_to_STOP...')
if keystroke == "":
    remove_temp_csv_files(output_path)
else:
    exit()
else:
    print('no_files_in:_)
    print(output_path)
    print('continue_execution.')
# for all subjects
for SUBJECT in range(0, NumberOfSubjects):
    print('**Subject_number', SUBJECT)
    # print('writing in file...')
    f = open('AllScoreSubject%02d.csv' % SUBJECT, 'w')
    # for a subjects each run, there are 6 runs in total
    for RUN in range(1, 6):
        print('Run_Number', RUN)
        # input files stored as DATA/Participant_%02d/Run_%d/log/EuroCab.log
        speed_curve_score('D:/GoogleDrive/1_OFc_Work/work/1-2_Thesis/code_and_data/Chapter4_exp/DATA/OPSIMU/'
                          'DATA/Participant_%02d/Run_%d/log/EuroCab.log' % (SUBJECT, RUN),
                          output_path + '/AllScoreSubject%02d.csv' % SUBJECT, RUN, SUBJECT)
    f.close()
    print('subject_number', SUBJECT, 'OK.')
print('csv_files_generated_in...')
print(output_path)
print('MERGE_MANUALLY._launch_CMD_and_execute:')
print('copy_*.csv_mergedAllSubjects.csv')
print('ATTENTION:_re--running_this_script_with_delete_ALL_csv_files_in_/SCOREALL_v2/_--_including_merged_file!!')

print('Start_execution...')

main()

```

A.2.4 Subjective questionnaires data - Pre, post questionnaire

The next page include the data collected from the pre-post questionnaires from the experimentation, [Equation 6.2.3.2](#). The data is grouped by the three groups that were created from the classification scores. The questions are presented in the horizontal tab. The rest of this page is left blank intentionally.

GROUP 1		pre_subj_01	pre_subj_07	pre_subj_09	pre_subj_10	pre_subj_11	pre_subj_13
Pre-Questionnaire / Participant ID:		Most of the things	Most of the things	Most of the things	yes, completely	yes, completely	yes, completely
PRE1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
PRE2. Do you understand your main objectives of this experimentation ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, it takes me time to adjust	No, it takes me time to adjust
PRE3. In general, are you able to understand a new situation quickly and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	I am not sure	No, it takes me time to adjust	Yes, most of the times
PRE4. Are you able to do multiple tasks (multitask) in a given time?		Yes	I am not sure	No, it takes me time to adjust	I am not sure	No, it takes me time to adjust	Yes, most of the times
PRE5. Are you physically tired?		Yes	I am not sure	No, I am completely attentive	I am not sure	No, I am completely attentive	Maybe
PRE6. Do you think you can perform all the scenarios that will be given to you successfully (based on the explanation...)?		Yes	Yes	Maybe	Yes	Yes	Yes
PRE7. Are you easily get frustrated?		no, I will try next time	no, I will try next time	depends	depends	depends	depends
Post-Questionnaire / Participant ID:		yes, completely	yes, completely	yes, completely	yes, completely	yes, completely	yes, completely
POST1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
POST2. Do you think you understood main objectives, what you should do ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, it took me time to adjust
POST3. Were you able to understand new situations (signals, tasks, etc.) and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	No, few times only	Yes, most of the times	No, few times only	No, it took me time to adjust
POST4. Were you able to multitask in the given scenario?		Yes	No	No	Yes	No	I am not sure
POST5. Are you physically tired?		No	Yes	No	No	No	Yes
POST6. Do you think you performed most of the tasks in scenarios successfully?		Yes	Yes	Yes	Yes	Yes	Yes
POST7. Did you put all your effort in the tasks?		no, I might have tried harder	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks
POST8. Did you get frustrated?		no, very few times	equally frustrated and not frustrated	no, very few times	no, very few times	yes, for most of the scenario runs	equally frustrated and not frustrated
GROUP 2		pre_subj_02	pre_subj_03	pre_subj_12	pre_subj_10	pre_subj_11	pre_subj_13
Pre-Questionnaire / Participant ID:		Most of the things	Most of the things	Most of the things	yes, completely	yes, completely	yes, completely
PRE1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
PRE2. Do you understand your main objectives of this experimentation ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, it took me time to adjust
PRE3. In general, are you able to understand a new situation quickly and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, few times only	No, it took me time to adjust
PRE4. Are you able to do multiple tasks (multitask) in a given time?		Yes	No	No	Yes	No	I am not sure
PRE5. Are you physically tired?		No	Yes	No	Yes	No	Yes
PRE6. Do you think you can perform all the scenarios that will be given to you successfully (based on the explanation...)?		Yes	Yes	Yes	Yes	Yes	Yes
PRE7. Are you easily get frustrated?		I am not sure	Yes	Yes, if I cannot do what I am supposed to	Yes	Yes	Yes
Post-Questionnaire / Participant ID:		Most of the things	Most of the things	Most of the things	yes, completely	yes, completely	yes, completely
POST1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
POST2. Do you think you understood main objectives, what you should do ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, it took me time to adjust
POST3. Were you able to understand new situations (signals, tasks, etc.) and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	No, few times only	No, it took me time to adjust
POST4. Were you able to multitask in the given scenario?		Yes	No	No	Yes	No	I am not sure
POST5. Are you physically tired?		No	Yes	No	Yes	No	Yes
POST6. Do you think you performed most of the tasks in scenarios successfully?		Yes	Yes	Yes	Yes	Yes	Yes
POST7. Did you put all your effort in the tasks?		no, I might have tried harder	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks
POST8. Did you get frustrated?		no, very few times	equally frustrated and not frustrated	no, very few times	no, very few times	yes, for most of the scenario runs	equally frustrated and not frustrated
GROUP 3		pre_subj_02	pre_subj_03	pre_subj_12	pre_subj_05	pre_subj_06	pre_subj_06
Pre-Questionnaire / Participant ID:		Most of the things	Most of the things	Most of the things	yes, completely	yes, completely	yes, completely
PRE1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
PRE2. Do you understand your main objectives of this experimentation ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	I am not sure	I am not sure
PRE3. In general, are you able to understand a new situation quickly and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	I am not sure	I am not sure
PRE4. Are you able to do multiple tasks (multitask) in a given time?		Yes	No	No	Yes	Yes	Yes
PRE5. Are you physically tired?		No	Yes	No	Yes	Yes	Yes
PRE6. Do you think you can perform all the scenarios that will be given to you successfully (based on the explanation...)?		Yes	Yes	Yes	Yes	Yes	Yes
PRE7. Are you easily get frustrated?		no, I will try next time	no, I will try next time	depends	depends	no, I will try next time	no, I will try next time
Post-Questionnaire / Participant ID:		Most of the things	Most of the things	Most of the things	yes, completely	yes, completely	yes, completely
POST1. Did you understand the explanations given to you?		Yes	Yes	Yes	Yes	Yes	Yes
POST2. Do you think you understood main objectives, what you should do ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	I am not sure	I am not sure
POST3. Were you able to understand new situations (signals, tasks, etc.) and be aware of what is happening ?		Yes, most of the times	Yes, most of the times	Yes, most of the times	Yes, most of the times	I am not sure	I am not sure
POST4. Were you able to multitask in the given scenario?		Yes	No	No	Yes	Yes	Yes
POST5. Are you physically tired?		No	Yes	No	Yes	Yes	Yes
POST6. Do you think you performed most of the tasks in scenarios successfully?		Yes	Yes	Yes	Yes	Yes	Yes
POST7. Did you put all your effort in the tasks?		no, I might have tried harder	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks	yes, fully in most of the tasks
POST8. Did you get frustrated?		no, very few times	equally frustrated and not frustrated	no, very few times	no, very few times	yes, for most of the scenario runs	equally frustrated and not frustrated

References

- [Aguirre et al., 2013a] Aguirre, F., Sallak, M., Schön, W., and Belmonte, F. (2013a). Application of evidential networks in quantitative analysis of railway accidents. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 227:368–384, doi: <http://dx.doi.org/10.1177/1748006X12475044>.
- [Aguirre et al., 2013b] Aguirre, F., Sallak, M., Vanderhaegen, F., and Berdjag, D. (2013b). An evidential network approach to support uncertain multiviewpoint abductive reasoning. *Information Sciences*, 253:110–125, doi: <http://dx.doi.org/10.1016/j.ins.2013.07.014>.
- [Almond, 1995] Almond, R. G. (1995). *Graphical Belief Modeling*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition.
- [Amalberti, 2001] Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, 37:109–126, doi: [http://dx.doi.org/10.1016/S0925-7535\(00\)00045-X](http://dx.doi.org/10.1016/S0925-7535(00)00045-X).
- [Aven, 2011] Aven, T. (2011). Interpretations of alternative uncertainty representations in a reliability and risk analysis context. *Reliability Engineering and System Safety*, 96(3):353–360, doi: <http://dx.doi.org/10.1016/j.ress.2010.11.004>.
- [Aven and Zio, 2011] Aven, T. and Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering and System Safety*, 96(1):64–74, doi: <http://dx.doi.org/10.1016/j.ress.2010.06.001>.
- [Avizienis et al., 2004] Avizienis, A., Laprie, J.-C., Randell, B., and Landwehr, C. (2004). Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1):11–33, doi: <http://dx.doi.org/10.1109/TDSC.2004.2>.
- [Bareith and Karsa, 2009] Bareith, A. and Karsa, Z. (2009). Simulator data collection for HRA purpose at NPP PAKS, Hungary. In *Simulator Data for HRA Purposes, OECDNEA Workshop, November*, pages 4–6.
- [Barnes et al., 2000] Barnes, M. J., Bley, D., and Cooper, S. (2000). Technical basis and implementation guidelines for a technique for human event analysis (ATHEANA). *NUREG-1624, Rev*, doi: <http://dx.doi.org/NUREG/CR-1624>.
- [Barot et al., 2013] Barot, V., Henshaw, M., Siemieniuch, C., Sinclair, M., Lim, S. L., Henson, S., Jamshidi, M., and DeLaurentis, D. (2013). State of the Art on Systems of Systems Management and Engineering. Technical report, e Trans-Atlantic Research and Education Agenda in Systems of Systems (T-AREA-SoS).

-
- [Barry, 1997] Barry, K. (1997). Validation of human reliability assessment techniques: Part 1 — Validation issues. *Safety Science*, 27(1):25–41, doi: [http://dx.doi.org/10.1016/S0925-7535\(97\)00049-0](http://dx.doi.org/10.1016/S0925-7535(97)00049-0).
- [Baysari et al., 2011] Baysari, M. T., Caponecchia, C., and McIntosh, A. S. (2011). A reliability and usability study of TRACER-RAV: The technique for the retrospective analysis of cognitive errors - For rail, Australian version. *Applied Ergonomics*, 42(6):852–859, doi: <http://dx.doi.org/10.1016/j.apergo.2011.01.009>.
- [Belmonte et al., 2011] Belmonte, F., Schön, W., Heurley, L., and Capel, R. (2011). Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway trafficsupervision. *Reliability Engineering and System Safety*, 96(2):237–249, doi: <http://dx.doi.org/10.1016/j.ress.2010.09.006>.
- [Bieder et al., 1998] Bieder, C., Le-Bot, P., Desmares, E., Bonnet, J. L., and Cara, F. (1998). MERMOS: EDF's new advanced HRA method.
- [Birolini, 2014] Birolini, A. (2014). *Reliability Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Blackman and Gertman, 1994] Blackman, H. S. and Gertman, D. I. (1994). *Human reliability and safety analysis data handbook*. John Wiley & Sons.
- [Blackman et al., 2008] Blackman, H. S., Gertman, D. I., and Boring, R. L. (2008). Human Error Quantification Using Performance Shaping Factors in the SPAR-H Method.
- [Boring, 2007] Boring, R. L. (2007). Dynamic Human Reliability Analysis: Benefits and Challenges of Simulating Human Performance. In *Proceedings of the European Safety and Reliability Conference (ESREL 2007)*.
- [Boring, 2010] Boring, R. L. (2010). How Many Performance Shaping Factors are Necessary for Human Reliability Analysis ? In *Proceedings of the 10th International Probabilistic Safety Assessment & Management Conference (PSAM10)*, pages 1479–1487.
- [Boring, 2012] Boring, R. L. (2012). Fifty Years of THERP and Human Reliability Analysis. In *Proceedings of the Probabilistic Safety Assessment and Management and European Safety and Reliability Conference (PSAM 11 & ESREL 2012)*. 2012, pages 3523–3532.
- [Boring, 2015] Boring, R. L. (2015). Defining Human Failure Events for Petroleum Applications of Human Reliability Analysis. In *AHFE Conference 2015, Procedia Manufacturing*, volume 3, pages 1335–1342. Procedia Manufacturing.
- [Boring and Blackman, 2007] Boring, R. L. and Blackman, H. S. (2007). The origins of the SPAR-H method's performance shaping factor multipliers. *IEEE Conference on Human Factors and Power Plants*, pages 177–184, doi: <http://dx.doi.org/10.1109/HFPP.2007.4413202>.
- [Boring et al., 2007] Boring, R. L., Griffith, C. D., and Joe, J. C. (2007). The measure of human error: Direct and indirect performance shaping factors. In *IEEE Conference on Human Factors and Power Plants*, pages 170–176.
- [Bot, 2010] Bot, P. L. (2010). Overview of the MERMOS Human Reliability Analysis method. *3rd International Symposium on Resilient Control Systems*, (August).
- [Bowie et al., 2015] Bowie, J., Munley, G., Dang, V., Wreathall, J., Bye, A., Cooper, S., Marble, J., Peters, S., Xing, J., Fauchille, V., Fiset, J. Y., Haage, M., Johanson, G., Jung, W. D., Kim, J., Lee,

- S. J., and Kubicek, J. A. (2015). Establishing the Appropriate Attributes in Current Human Reliability Assessment Techniques for Nuclear Safety (NEA-CSNI-R-2015-1). Technical Report March, Nuclear Energy Agency of the OECD (NEA).
- [Budnitz et al., 1997] Budnitz, R. J., Apostolakis, G., Boore, D. M., Cluff, L. S., Coppersmith, K. J., Cornell, C. a., and Morris, P. a. (1997). Recommendations for Probabilistic Seismic Hazard Analysis : Guidance on Uncertainty and Use of Experts. *Power*, 1:998–1006, doi: <http://dx.doi.org/NUREG/CR-6372Vol1.1>.
- [Bye et al., 2016] Bye, A., Laumann, K., Taylor, C., Rasmussen, M., Øie, S., van de Merwe, K., Øien, K., Boring, R. L., Paltrinieri, N., Wærø, I., Massaiu, S., and Gould, K. (2016). Petro-hra, a New Method for Human Reliability Analysis in the Petroleum Industry. *Proceedings of the International Conference on Probabilistic Safety Assessment and Management (PSAM 13) 2-7 October, 2016, Seoul, Korea*, (October).
- [CENELEC, 1999] CENELEC (1999). 50126 (1999): Railway applications–The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS). Technical report, Availability, Maintainability and Safety (RAMS) (1999).
- [Chandler et al., 2006a] Chandler, T., Chang, J., Mosleb, A., J., M., Boring, R., and Gertman, D. (2006a). Human Reliability Analysis Methods Selection Guidance for NASA. *National Aeronautics and Space Administration*, (July):175.
- [Chandler et al., 2006b] Chandler, T., Chang, J., Mosleb, A., J., M., Boring, R., and Gertman, D. (2006b). Human Reliability Analysis Methods Selection Guidance for NASA. Technical Report July, National Aeronautics and Space Administration. url: <http://www.hq.nasa.gov/office/codeq/rm/docs/HRA{ }Report.pdf>.
- [Comisión de investigación de accidentes ferroviarios, 2014] Comisión de investigación de accidentes ferroviarios (2014). Informe final sobre el accidente grave ferroviario nº 0054/2013 ocurrido el día 24.07.2013 en las proximidades de la estación de Santiago de Compostela (A Coruña). Technical report, COMISIÓN DE INVESTIGACIÓN DE ACCIDENTES FERROVIARIOS. url: <http://www.fomento.gob.es/NR/rdonlyres/0ADE7F17-84BB-4CBD-9451-C750EDE06170/125127/IF240713200514CIAF.pdf>.
- [Connelly et al., 2012] Connelly, S., Hussey, A., and Becht, H. (2012). Practical Early-lifecycle Application of Human Factors Assessment. In *Proceedings of the Australian System Safety Conference - Volume 145, ASSC '12*, pages 47–54, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.
- [Cooper et al., 1996] Cooper, S. E., Ramey-Smith, A. M., Wreathall, J., Parry, G. W., Bley, D. C., Luckas, W. J., Taylor, J. H., and Barriere, M. T. (1996). A Technique for Human Error Analysis (ATHEANA). Technical report, U.S. Nuclear Regulatory Commission, Brookhaven National Laboratory, Upton.
- [Cuzzolin, 2008] Cuzzolin, F. (2008). A geometric approach to the theory of evidence. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 38(4):522–534.
- [De Galizia et al., 2016a] De Galizia, A., Simon, C., Weber, P., Iung, B., Duval, C., and Serdet, E. (2016a). Modelling Non-Deterministic Causal Mechanisms involving Resilience in Risk Analysis. *IFAC-PapersOnLine*, 49(12):325–330, doi: <http://dx.doi.org/10.1016/j.ifacol.2016.07.625>.

-
- [De Galizia et al., 2016b] De Galizia, A., Simon, C., Weber, P., Iung, B., Duval, C., and Serdet, E. (2016b). Modelling Non-Deterministic Causal Mechanisms involving Resilience in Risk Analysis. *IFAC-PapersOnLine*, 49(12):325–330, doi: <http://dx.doi.org/10.1016/j.ifacol.2016.07.625>.
- [Dempster, 1967] Dempster, A. P. (1967). Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, doi: <http://dx.doi.org/10.2307/2239146>.
- [Det Norske Veritas, 2010] Det Norske Veritas (2010). Final Report – Risk Acceptance Criteria for Technical Systems and Operational Procedures. *European Railway Agency*, (January).
- [Di Grazia et al., 2014] Di Grazia, G., Vittorini, B., Carlizza, L., Lamedica, R., and Fabbri, G. (2014). Human factor relevance in safety assurance of railway operations. In *AET Annual Conference*, pages 1–6. IEEE.
- [Di Pasquale et al., 2013] Di Pasquale, V., Iannone, R., Miranda, S., and Riemma, S. (2013). An Overview of Human Reliability Analysis Techniques in Manufacturing Operations. *Operations Management*, pages 221–240.
- [Dougherty, 1990] Dougherty, E. (1990). Human reliability analysis?where shouldst thou turn? *Reliability Engineering & System Safety*, 29(3):283–299, doi: [http://dx.doi.org/10.1016/0951-8320\(90\)90012-C](http://dx.doi.org/10.1016/0951-8320(90)90012-C).
- [Duval et al., 2012] Duval, C., Fallet-Fidry, G., Iung, B., Weber, P., and Levrat, E. (2012). A Bayesian network-based integrated risk analysis approach for industrial systems: application to heat sink system and prospects development. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 226(5):488–507, doi: <http://dx.doi.org/10.1177/1748006X12451091>.
- [Edkins and Pollock, 1997] Edkins, G. D. and Pollock, C. M. (1997). The influence of sustained attention on Railway accidents. *Accident Analysis & Prevention*, 29(4):533–539, doi: [http://dx.doi.org/10.1016/S0001-4575\(97\)00033-X](http://dx.doi.org/10.1016/S0001-4575(97)00033-X).
- [Embrey, 1986] Embrey, D. E. (1986). SHERPA: A systematic human error reduction and prediction approach. In *Paper Presented at the International Topical Meeting on Advances in Human Factors in Nuclear Power Systems*, Knoxville, Tennessee.
- [Embrey et al., 1984] Embrey, D. E., Humphreys, P., Rosa, E. A., Kirwan, B., and Rea, K. (1984). SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment, Volume 2: Detailed Analysis of the Technical Issues NUREG/CR-3518. Technical report, Brookhaven National Laboratory, Brookhaven National Laboratory, Upton.
- [Endsley, 1995] Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, doi: <http://dx.doi.org/10.1518/001872095779049543>.
- [European Commission, 2011] European Commission (2011). MODSafe Risk Analysis. pages 1–12.
- [European Parliament, 2004] European Parliament (2004). DIRECTIVE 2004/49/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2004 on safety on the Community's railways and amending Council Directive 95/18/EC on the licensing of railway undertakings and Directive 2001/14/EC on the allocation of railwa.

-
- [European Railway Agency, 2009] European Railway Agency (2009). Collection of examples of risk assessments and of some possible tools supporting the CSM Regulation. 33(0):1–105.
- [European Railway Agency, 2015a] European Railway Agency (2015a). Accompanying Report N. ERA-REC-116-2015-ACR to the Recommendation of the European Railway Agency on the amendment of the Commission implementing Regulation (EU) No 402/2013 on the common safety method for risk evaluation and assessment. Technical report, European Railway Agency.
- [European Railway Agency, 2015b] European Railway Agency (2015b). Recommendation of the European Railway Agency on the amendment of the Commission implementing Regulation (EU) No 402/2013 on the common safety method for risk evaluation and assessment. Technical report, European Railway Agency.
- [European Railway Agency, 2016] European Railway Agency (2016). Introduction to ETCS Braking curves. Reference: ERA_ERTMS_040026. Technical report. url: papers3://publication/uuid/65EDA06E-FF69-4FFF-BBC7-247678A5DF1B.
- [Evans, 2011] Evans, A. W. (2011). Fatal train accidents on Europe's railways: 1980-2009. *Accident Analysis and Prevention*, 43(1):391–401, doi: <http://dx.doi.org/10.1016/j.aap.2010.09.009>.
- [Feldmann et al., 2008] Feldmann, F., Hammerl, M., and Schwartz, S. (2008). Questioning human error probabilities in railways. In *3rd IET International Conference on System Safety 2008*, pages 4B2–4B2. IEE.
- [Forester et al., 2004] Forester, J., Bley, D., Cooper, S., Lois, E., Siu, N., Kolaczowski, A., and Wreathall, J. (2004). Expert elicitation approach for performing ATHEANA quantification. *Reliability Engineering and System Safety*, 83(2):207–220, doi: <http://dx.doi.org/10.1016/j.res.2003.09.011>.
- [French et al., 2011] French, S., Bedford, T., Pollard, S. J., and Soane, E. (2011). Human reliability analysis: A critique and review for managers. *Safety Science*, 49(6):753–763, doi: <http://dx.doi.org/10.1016/j.ssci.2011.02.008>.
- [Galizia et al., 2015] Galizia, A. D., Duval, C., Serdet, E., Weber, P., Simon, C., Galizia, A. D., Duval, C., Serdet, E., Weber, P., Simon, C., De Galizia, A., Duval, C., Serdet, E., Weber, P., Simon, C., and Iung, B. (2015). Advanced investigation of HRA methods for probabilistic assessment of human barriers efficiency in complex systems for a given organisational and environmental context. In *International Topical Meeting on Probabilistic Safety Assessment and Analysis, PSA 2015*, Sun Valley, United States. Idaho National Laboratory/Idaho American Nuclear Society.
- [Gaur, 2005] Gaur, D. (2005). Human Factors Analysis and Classification System applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine*, 76(Number 5):501–505(5).
- [Gertman et al., 2005] Gertman, D., Blackman, H., Marble, J., Byers, J., and Smith, C. (2005). The SPAR-H human reliability analysis method. Technical report, Idaho National Laboratory U.S. url: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6883/>.
- [Giang and Shenoy, 2003] Giang, P. H. and Shenoy, S. (2003). The Belief Function Machine (BFM) An Environment for Reasoning with Belief Functions in MatLab. Technical report, Univ. of Kansas. url: <http://pshenoy.faculty.ku.edu/Papers/BFM072503.zip>.

-
- [Gibson et al., 2013] Gibson, H. W., Mills, A., Smith, S., and Kirwan, B. K. (2013). Railway Action Reliability Assessment A Railway - Specific Approach to Human Error Quantification. *Rail Human Factors Supporting Reliability, Safety and Cost Reduction*, pages 671–676.
- [Groth, 2009] Groth, K. M. . (2009). *A data-informed model of performance shaping factors for use in human reliability analysis*. PhD thesis, University of Maryland, College Park.
- [Groth and Mosleh, 2012a] Groth, K. M. and Mosleh, A. (2012a). A data-informed PIF hierarchy for model-based Human Reliability Analysis. *Reliability Engineering and System Safety*, 108:154–174, doi: <http://dx.doi.org/10.1016/j.ress.2012.08.006>.
- [Groth and Mosleh, 2012b] Groth, K. M. and Mosleh, A. (2012b). Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 226(4):361–379, doi: <http://dx.doi.org/10.1177/1748006X11428107>.
- [Groth et al., 2014] Groth, K. M., Smith, C. L., and Swiler, L. P. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. *Reliability Engineering & System Safety*, 128:32–40, doi: <http://dx.doi.org/10.1016/j.ress.2014.03.010>.
- [Groth and Swiler, 2013] Groth, K. M. and Swiler, L. P. (2013). Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H. *Reliability Engineering & System Safety*, 115:33–42, doi: <http://dx.doi.org/10.1016/j.ress.2013.02.015>.
- [Ha and Seong, 2009] Ha, J. S. and Seong, P. H. (2009). HUPESS: Human Performance Evaluation Support System. *Reliability and Risk Issues in Large Scale Safety-critical Digital Control Systems*, pages 197–229.
- [Hallbert et al., 2012] Hallbert, B. P., Morgan, T., Hugo, J., Oxstrand, J., and Persensky, J. (2012). A Formalized Approach for the Collection of HRA Data from Nuclear Power Plant Simulators. (MAY 2014).
- [Hammerl and Vanderhaegen, 2012] Hammerl, M. and Vanderhaegen, F. (2012). Human factors in the railway system safety analysis process. In *Rail Human Factors Around the World: Impacts on and of People for Successful Rail Operations: 3rd International Rail Human Factors Conference*, volume 1, pages 73–84.
- [Hammitt and Zhang, 2013] Hammitt, J. K. and Zhang, Y. (2013). Combining Experts' Judgments: Comparison of Algorithmic Methods Using Synthetic Data. *Risk Analysis*, 33(1):109–120, doi: <http://dx.doi.org/10.1111/j.1539-6924.2012.01833.x>.
- [Hankins and Wilson, 1998] Hankins, T. C. and Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine*, 69(4):360–367.
- [Hart and Staveland, 1988] Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52(C):139–183, doi: [http://dx.doi.org/10.1016/S0166-4115\(08\)62386-9](http://dx.doi.org/10.1016/S0166-4115(08)62386-9).
- [Hart, Sandra, 2006] Hart, Sandra, G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Human Factors and Ergonomics Society Annual Meeting*, pages 904–908, doi: <http://dx.doi.org/10.1037/e577632012-009>.

-
- [Health and Safety Executive, 2009] Health and Safety Executive (2009). Review of human reliability assessment methods. Technical report. url: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Review+of+human+reliability+assessment+methods{#}0>.
- [Henson et al., 2013] Henson, S., Henshaw, M., Barot, V., Siemieniuch, C., Sinclair, M., Jamshidi, M., Dogan, H., Lim, S., Ncube, C., and DeLaurentis, D. (2013). Towards a Systems of Systems Engineering EU Strategic Research Agenda. *2013 8th International Conference on System of Systems Engineering*, pages 99–104, doi: <http://dx.doi.org/10.1109/SYSSE.2013.6575250>.
- [Hill et al., 1992] Hill, S., Lavecchia, H., Byers, J., Bittner, A., Zaklad, A., and Christ, R. (1992). Comparison of four subjective workload rating scales. *Human Factors*, 34(4):429–439, doi: <http://dx.doi.org/10.1177/001872089203400405>.
- [Hollnagel, 1998] Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Elsevier Science Ltd.
- [Høyland and Rausand, 1994] Høyland, a. and Rausand, M. (1994). *System reliability theory: models and statistical methods*.
- [J. Forester et al., 2007] J. Forester, Kolaczowski, A., Cooper, S., Bley, D., Lois, E., Forester, J., Kolaczowski, A., Cooper, S., Bley, D., and Lois, E. (2007). ATHEANA User’s Guide Final Report NUREG-1880. Technical report, U.S. Nuclear Regulatory Commission. url: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1880>.
- [James Chang et al., 2014] James Chang, Y., Bley, D., Criscione, L., Kirwan, B., Mosleh, A., Madary, T., Nowell, R., Richards, R., Roth, E. M., Sieben, S., and Zouli, A. (2014). The SACADA database for human reliability and human performance. *Reliability Engineering and System Safety*, 125:117–133, doi: <http://dx.doi.org/10.1016/j.ress.2013.07.014>.
- [Jamshidi, 2011] Jamshidi, M. (2011). *Systems of Systems Engineering: Principles and Applications*, volume 36.
- [Johnson, 2006] Johnson, C. W. (2006). What are emergent properties and how do they affect the engineering of complex systems? *Reliability Engineering and System Safety*, 91(12):1475–1481, doi: <http://dx.doi.org/10.1016/j.ress.2006.01.008>.
- [Jousselme et al., 2001] Jousselme, A.-L., Grenier, D., and Bossé, É. (2001). A new distance between two bodies of evidence. *Information Fusion*, 2(2):91–101, doi: [http://dx.doi.org/10.1016/S1566-2535\(01\)00026-4](http://dx.doi.org/10.1016/S1566-2535(01)00026-4).
- [Kecklund et al., 2013] Kecklund, L., Mowitz, A., and Antova, M. (2013). Current practices of the assessment and acceptance of risks related to human interactions within the european railways. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, (January):508–516.
- [KIM and BISHU, 2006] KIM, B. J. and BISHU, R. R. (2006). UNCERTAINTY OF HUMAN ERROR AND FUZZY APPROACH TO HUMAN RELIABILITY ANALYSIS.
- [Kim and Jung, 2003] Kim, J. W. and Jung, W. (2003). A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries*, 16(6):479–495, doi: [http://dx.doi.org/10.1016/S0950-4230\(03\)00075-5](http://dx.doi.org/10.1016/S0950-4230(03)00075-5).
- [Kim et al., 2006] Kim, M. C., Seong, P. H., and Hollnagel, E. (2006). A probabilistic approach for determining the control mode in CREAM. *Reliability Engineering and System Safety*, 91(2):191–199, doi: <http://dx.doi.org/10.1016/j.ress.2004.12.003>.

-
- [Kinder et al., 2015] Kinder, A., Henshaw, M., Siemieniuch, C., and Wrigley, C. (2015). A Model Based Approach to System of Systems Risk Management. In *10th System of Systems Engineering Conference (SoSE)*, pages 122–127.
- [Kirwan and Gibson, 2007] Kirwan, B. and Gibson, H. (2007). CARA: A Human Reliability Assessment Tool for Air Traffic Safety Management — Technical Basis and Preliminary Architecture. *The Safety of Systems*, pages 197–214, doi: http://dx.doi.org/10.1007/978-1-84628-806-7_13.
- [Kirwan et al., 2004] Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G., and Umbers, I. (2004). Nuclear Action Reliability Assessment (NARA): A Data-Based HRA Tool. In *Probabilistic Safety Assessment and Management*, pages 1206–1211. Springer London, London.
- [Knol et al., 2008] Knol, A., Sluijs, J. P. V. D., and Slottje, P. (2008). Expert Elicitation : Methodological suggestions for its use in environmental health impact assessments. *RIVM Letter report 630004001/2008*, page 56.
- [Kolaczowski et al., 2005] Kolaczowski, A., Forester, J., Lois, E., and Cooper, S. (2005). Good Practices for Implementing Human Reliability Analysis (HRA) (NUREG-1792). Technical report, U.S. Nuclear Regulatory Commission. url: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1792/>.
- [Kyriakidis, 2013] Kyriakidis, M. (2013). Developing a human performance railway operational index to enhance safety of railway operations. *Imperial College London*, (October 2013).
- [Kyriakidis et al., 2011] Kyriakidis, M., Majumdar, A., Grote, G., and Ochieng, W. Y. (2011). Development and Assessment of Taxonomy for Performance-Shaping Factors for Railway Operations. *SSRN Electronic Journal*, 2289(-1):145–153, doi: <http://dx.doi.org/10.2139/ssrn.2506591>.
- [Kyriakidis et al., 2012] Kyriakidis, M., Majumdar, A., and Ochieng, W. Y. (2012). A human performance operational railway index to estimate operator’s error probability. *Advances in Human Aspects of Road and Rail Transportation*, pages 832–841, doi: <http://dx.doi.org/10.2139/ssrn.2506631>.
- [Kyriakidis et al., 2015a] Kyriakidis, M., Majumdar, A., and Ochieng, W. Y. (2015a). Data based framework to identify the most significant performance shaping factors in railway operations. *Safety Science*, 78:60–76, doi: <http://dx.doi.org/10.1016/j.ssci.2015.04.010>.
- [Kyriakidis et al., 2015b] Kyriakidis, M., Pak, K. T., and Majumdar, A. (2015b). Railway Accidents due to Human Error A historic analysis of the UK railways (1945-2012). *Transportation Research Record: Journal of the Transportation Research Board*, pages 1–16.
- [Langseth and Portinale, 2007] Langseth, H. and Portinale, L. (2007). Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, doi: <http://dx.doi.org/10.1016/j.res.2005.11.037>.
- [Laprie, 1992] Laprie, J.-C. (1992). *Dependability: Basic Concepts and Terminology*, volume 5 of *Dependable Computing and Fault-Tolerant Systems*. Springer Vienna, Vienna.
- [Le Bot, 2004] Le Bot, P. (2004). Human reliability data, human error and accident models—illustration through the Three Mile Island accident analysis. *Reliability Engineering & System Safety*, 83(2):153–167, doi: <http://dx.doi.org/10.1016/j.res.2003.09.007>.

-
- [Leveson, 2015] Leveson, N. (2015). A systems approach to risk management through leading safety indicators. *Reliability Engineering and System Safety*, 136:17–34, doi: <http://dx.doi.org/10.1016/j.ress.2014.10.008>.
- [Leveson, 2011] Leveson, N. G. (2011). Applying systems thinking to analyze and learn from events. *Safety Science*, 49(1):55–64, doi: <http://dx.doi.org/10.1016/j.ssci.2009.12.021>.
- [Lois et al., 2009] Lois, E., Dang, V. N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P. O., Parry, G., Julius, J., Boring, R., Männistö, I., and Bye, A. (2009). International HRA Empirical Study – Phase 1 Report (NUREG/IA-0216, Volume 1). Technical Report November 2009, Office of Nuclear Regulatory Research U.S. Nuclear Regulatory Commission.
- [Loudahi et al., 2014] Loudahi, M., Klein, J., Vannobel, J. M., and Colot, O. (2014). New distances between bodies of evidence based on Dempsterian specialization matrices and their consistency with the conjunctive combination rule. *International Journal of Approximate Reasoning*, 55(5):1093–1112, doi: <http://dx.doi.org/10.1016/j.ijar.2014.02.007>.
- [Lyons et al., 2004] Lyons, M., Woloshynowych, M., Adams, S., and Vincent, C. (2004). Error Reduction in Medicine. page 81.
- [Marseguerra et al., 2007] Marseguerra, M., Zio, E., and Librizzi, M. (2007). Human reliability analysis by fuzzy "CREAM". *Risk Analysis*, 27(1):137–154, doi: <http://dx.doi.org/10.1111/j.1539-6924.2006.00865.x>.
- [Martins and Maturana, 2013] Martins, M. R. and Maturana, M. C. (2013). Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliability Engineering and System Safety*, 110:89–109, doi: <http://dx.doi.org/10.1016/j.ress.2012.09.008>.
- [McLeod et al., 2007] McLeod, R. W., Walker, G. H., and Mills, a. (2007). Assessing the human factors risks in extending the use of AWS.
- [Meyer et al., 2007] Meyer, P., Le Bot, P., and Pesme, H. (2007). MERMOS: An extended second generation HRA method. *IEEE Conference on Human Factors and Power Plants*, pages 276–283, doi: <http://dx.doi.org/10.1109/HFPP.2007.4413219>.
- [Miller, 2001] Miller, S. (2001). Literature Review - Workload measures. Technical Report August.
- [Mkrtchyan et al., 2015] Mkrtchyan, L., Podofillini, L., and Dang, V. (2015). Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability Engineering & System Safety*, 139:1–16, doi: <http://dx.doi.org/10.1016/j.ress.2015.02.006>.
- [Mkrtchyan et al., 2016] Mkrtchyan, L., Podofillini, L., and Dang, V. N. (2016). Methods for building Conditional Probability Tables of Bayesian Belief Networks from limited judgment: An evaluation for Human Reliability Application. *Reliability Engineering and System Safety*, 151:93–112, doi: <http://dx.doi.org/10.1016/j.ress.2016.01.004>.
- [Mosleh and Chang, 2004] Mosleh, A. and Chang, Y. H. (2004). Model-based human reliability analysis: Prospects and requirements. *Reliability Engineering and System Safety*, 83:241–253, doi: <http://dx.doi.org/10.1016/j.ress.2003.09.014>.
- [Mowitz and Kecklund, 2013] Mowitz, A. O. and Kecklund, L. (2013). Study on the Assessment and the Acceptance of Risks Related to Human Interactions within the European Railways ERA/2011/SAF/OP/02 Aino. Technical Report January. url: <http://www.era.europa.eu/Document-Register/Pages/Human-RAC-study.aspx>.

-
- [Musharraf et al., 2014] Musharraf, M., Bradbury-Squires, D., Khan, F., Veitch, B., Mackinnon, S., and Imtiaz, S. (2014). A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. *Reliability Engineering and System Safety*, 132:1–8, doi: <http://dx.doi.org/10.1016/j.res.2014.06.016>.
- [Ouchi, 2004] Ouchi, F. (2004). A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis. (February):20, doi: <http://dx.doi.org/10.1596/1813-9450-3201>.
- [Ouedraogo et al., 2013] Ouedraogo, K. A., Enjalbert, S., and Vanderhaegen, F. (2013). How to learn from the resilience of Human-Machine Systems? *Engineering Applications of Artificial Intelligence*, 26(1):24–34, doi: <http://dx.doi.org/10.1016/j.engappai.2012.03.007>.
- [Park and Jung, 2007] Park, J. and Jung, W. (2007). OPERA-a human performance database under simulated emergencies of nuclear power plants. *Reliability Engineering and System Safety*, 92(4):503–519, doi: <http://dx.doi.org/10.1016/j.res.2006.01.007>.
- [Parry, 1996] Parry, G. W. (1996). The characterization of uncertainty in probabilistic risk assessments of complex systems. *Reliability Engineering and System Safety*, 54(2-3):119–126, doi: [http://dx.doi.org/10.1016/S0951-8320\(96\)00069-5](http://dx.doi.org/10.1016/S0951-8320(96)00069-5).
- [Pearl, 2014] Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- [Perrow, 1999] Perrow, C. (1999). *Normal accidents: Living with high risk technologies*. Princeton University Press.
- [Pickup et al., 2013] Pickup, L., Ryan, B., Atkinson, S., Dadashi, N., Golightly, D., and Wilson, J. (2013). Support Study for Human Factors Integration – Human Functions in European Railways: Final Report for the European Railway Agency. Technical Report Final Report for the European Railway Agency ERA/2012/SAF/NP/02, ERA. url: <http://www.era.europa.eu/Document-Register/Pages/Study-Human-Factors-Integration.aspx>.
- [Podofillini and Dang, 2013] Podofillini, L. and Dang, V. (2013). A Bayesian approach to treat expert-elicited probabilities in human reliability analysis model construction. *Reliability Engineering & System Safety*, 117:52–64, doi: <http://dx.doi.org/10.1016/j.res.2013.03.015>.
- [Polet et al., 2003] Polet, P., Vanderhaegen, F., and Amalberti, R. (2003). Modelling border-line tolerated conditions of use (BTCU) and associated risks. *Safety Science*, 41(2-3):111–136, doi: [http://dx.doi.org/10.1016/S0925-7535\(02\)00037-1](http://dx.doi.org/10.1016/S0925-7535(02)00037-1).
- [Qiu, 2014] Qiu, S. (2014). *Graphical models for RAMS assessment and risk analysis of Systems of Systems under uncertainty*. PhD thesis, Université de technologie de Compiègne.
- [Qiu et al., 2017] Qiu, S., Rachedi, N., Sallak, M., and Vanderhaegen, F. (2017). A quantitative model for the risk evaluation of driver-ADAS systems under uncertainty. *Reliability Engineering & System Safety*, doi: <http://dx.doi.org/10.1016/j.res.2017.05.028>.
- [Qiu et al., 2015] Qiu, S., Sacile, R., Sallak, M., and Schön, W. (2015). On the application of Valuation-Based Systems in the assessment of the probability bounds of Hazardous Material transportation accidents occurrence. *Safety Science*, 72(Dm):83–96, doi: <http://dx.doi.org/10.1016/j.ssci.2014.08.006>.

-
- [Qiu et al., 2014] Qiu, S., Sallak, M., Schön, W., and Cherfi-boulanger, Z. (2014). Modeling of ERTMS Level 2 as an SoS and Evaluation of its Dependability Parameters Using Statecharts. *IEEE Systems Journal*, 8(4):1–13, doi: <http://dx.doi.org/10.1109/JSYST.2013.2297751>.
- [Qureshi, 2007] Qureshi, Z. H. (2007). A review of accident modelling approaches for complex socio-technical systems | University of Queensland. *12th Australian workshop on Safety critical systems and software and safety-related programmable systems*, 86:47–59.
- [Rachedi et al., 2012] Rachedi, N.-D.-E., Berdjag, D., and Vanderhaegen, F. (2012). Détection de l'état d'un opérateur humain dans le contexte de la conduite ferroviaire. In *8ème Congrès de Maitrise des Risques et Sureté de Fonctionnement*, number Ccm, pages 1–7.
- [Rail Safety & Standards Board, 2008] Rail Safety & Standards Board (2008). Understanding Human Factors a guide for the railway industry.
- [Rail Safety and Standards Board (RSSB), 2008] Rail Safety and Standards Board (RSSB) (2008). Good Practice Guide on Cognitive and Individual Risk Factors. (1):1–64.
- [Rangra et al., 2017a] Rangra, S., Sallak, M., Schön, W., and Vanderhaegen, F. (2017a). A graphical model based on performance shaping factors for assessing human reliability. *IEEE Transactions on Reliability*, PP(99):1–24, doi: <http://dx.doi.org/10.1109/TR.2017.2755543>.
- [Rangra et al., 2015a] Rangra, S., Sallak, M., Schön, W., and Vanderhaegen, F. (2015a). Human Reliability Assessment under Uncertainty – Towards a Formal Method. In *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015*, volume 3, pages 3230–3237. Elsevier B.V. Procedia Manufacturing.
- [Rangra et al., 2015b] Rangra, S., Sallak, M., Schön, W., and Vanderhaegen, F. (2015b). On the study of human reliability in transportation systems of systems. In *2015 10th System of Systems Engineering Conference (SoSE)*, pages 208–213, San Antonio, TX, USA. IEEE.
- [Rangra et al., 2017b] Rangra, S., Sallak, M., Schon, W., and Vanderhaegen, F. (2017b). A Graphical Model Based on Performance Shaping Factors for Assessing Human Reliability. *IEEE Transactions on Reliability*, 66(4):1120–1143, doi: <http://dx.doi.org/10.1109/TR.2017.2755543>.
- [Rasmussen, 1982] Rasmussen, J. (1982). Human errors. A taxonomy for describing human malfunction in industrial installations. *Journal of Occupational Accidents*, 4(2-4):311–333, doi: [http://dx.doi.org/10.1016/0376-6349\(82\)90041-4](http://dx.doi.org/10.1016/0376-6349(82)90041-4).
- [Rasmussen, 1997] Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, 27(2):183–213, doi: [http://dx.doi.org/10.1016/S0925-7535\(97\)00052-0](http://dx.doi.org/10.1016/S0925-7535(97)00052-0).
- [Reason, 1990] Reason, J. (1990). *Human Error*. Cambridge University Press.
- [Reason, 2000] Reason, J. (2000). Human error: models and management. *BMJ*, 320(7237):768–770, doi: <http://dx.doi.org/10.1136/bmj.320.7237.768>.
- [Reer, 2008] Reer, B. (2008). Review of advances in human reliability analysis of errors of commission-Part 2: EOC quantification. *Reliability Engineering and System Safety*, 93(8):1105–1122, doi: <http://dx.doi.org/10.1016/j.ress.2007.10.001>.
- [Sallak et al., 2013] Sallak, M., Schön, W., and Aguirre, F. (2013). Extended component importance measures considering aleatory and epistemic uncertainties. *IEEE Transactions on Reliability*, 62(1):49–65, doi: <http://dx.doi.org/10.1109/TR.2013.2240888>.

-
- [Salmon et al., 2005] Salmon, P., Regan, M., and Johnston, I. (2005). *Human Error and Road Transport*. Number 256.
- [Samad and Parisini, 2011] Samad, T. and Parisini, T. (2011). Systems of Systems. *IEEE Control Systems Society*.
- [Schön et al., 2013] Schön, W., Larrauffie, G., Moens, G., and Pore, J. (2013). *Railway Signalling and Automation Volume 1*. La Vie du Rail.
- [Schwencke et al., 2012] Schwencke, D., Lindner, T., Milius, B., Arenius, M., Sträter, O., and Lemmer, K. (2012). A new method for human reliability assessment in railway transport. *11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012, PSAM11 ESREL 2012*, 8(2):6139–6147.
- [Sebbak et al., 2014] Sebbak, F., Benhammadi, F., Mataoui, M., Bouznad, S., and Amirat, Y. (2014). An alternative combination rule for evidential reasoning. . . . (*FUSION*), 2014 17th
- [Sedki et al., 2013] Sedki, K., Polet, P., and Vanderhaegen, F. (2013). Using the BCD model for risk analysis: An influence diagram based approach. *Engineering Applications of Artificial Intelligence*, 26(9):2172–2183, doi: <http://dx.doi.org/10.1016/j.engappai.2013.06.009>.
- [Sentz and Ferson, 2002] Sentz, K. and Ferson, S. (2002). Combination of Evidence in Dempster-Shafer Theory.
- [Shafer, 1976] Shafer, G. (1976). *A mathematical theory of evidence*. Princeton: Princeton university press.
- [Sharek, 2009] Sharek, D. (2009). NASA-TLX Online Tool [desktop app].
- [Shenoy, 1989] Shenoy, P. P. (1989). A valuation-based language for expert systems. *International Journal of Approximate Reasoning*, 3(5):383–411, doi: [http://dx.doi.org/10.1016/0888-613X\(89\)90009-1](http://dx.doi.org/10.1016/0888-613X(89)90009-1).
- [Shenoy, 1992] Shenoy, P. P. (1992). Valuation-Based Systems for Bayesian Decision Analysis. *Operations Research*, 40(3):463–484, doi: <http://dx.doi.org/10.1287/opre.40.3.463>.
- [Shenoy, 1994] Shenoy, P. P. (1994). Using dempster-shafer’s belief-function theory in expert systems. *Advances in the Dempster-Shafer Theory of Evidence*, pages 395–414, doi: <http://dx.doi.org/10.1117/12.56867>.
- [Sheridan, 2008] Sheridan, T. B. (2008). Risk, Human Error, and System Resilience: Fundamental Ideas. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):418–426, doi: <http://dx.doi.org/10.1518/001872008X250773>.
- [Shirley et al., 2015] Shirley, R. B., Smidts, C., Li, M., and Gupta, A. (2015). Validating THERP: Assessing the scope of a full-scale validation of the Technique for Human Error Rate Prediction. *Annals of Nuclear Energy*, 77:194–211, doi: <http://dx.doi.org/10.1016/j.anucene.2014.10.017>.
- [Skjerve and Bye, 2011] Skjerve, A. B. and Bye, A. (2011). Simulator-based human factors studies across 25 years: The history of the Halden man-machine laboratory. *Simulator-based Human Factors Studies Across 25 Years: The History of the Halden Man-Machine Laboratory*, pages 1–364, doi: <http://dx.doi.org/10.1007/978-0-85729-003-8>.
- [Smets, 1992] Smets, P. (1992). The Nature of the unnormalized Beliefs encountered in the Transferable Belief Model. *Uncertainty in Artificial Intelligence*, (0):292–297, doi: <http://dx.doi.org/10.1016/B978-1-4832-8287-9.50044-X>.

-
- [Smets, 1993] Smets, P. (1993). Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem. *International Journal of Approximate Reasoning*, 9(1):1–35, doi: [http://dx.doi.org/10.1016/0888-613X\(93\)90005-X](http://dx.doi.org/10.1016/0888-613X(93)90005-X).
- [Smets, 1999] Smets, P. (1999). Practical Uses of Belief Functions. *Uncertainty in Artificial Intelligence 15. UAI99*, pages 612–621.
- [SNCF Réseau, 2016] SNCF Réseau (2016). Principes et règles d'exploitation du système ETCS - Document d'exploitation. Technical report.
- [Spurgin, 2009] Spurgin, A. J. (2009). *Human Reliability Assessment Theory and Practice*. CRC Press.
- [Stamatelatos et al., 2011] Stamatelatos, M., Dezfuli, H., Apostolakis, G., Everline, C., Guarro, S., Mathias, D., Mosleh, A., and Al., E. (2011). Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Technical Report December. url: <http://www.hq.nasa.gov/office/codeq/doctree/praguide.pdf>.
- [Su et al., 2015] Su, X., Mahadevan, S., Xu, P., and Deng, Y. (2015). Dependence Assessment in Human Reliability Analysis Using Evidence Theory and AHP. *Risk Analysis*, 35(7):1296–1316, doi: <http://dx.doi.org/10.1111/risa.12347>.
- [Swain and Guttman, 1983] Swain, A. and Guttman, H. (1983). Handbook of human-reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278. Technical report, Sandia National Labs., Albuquerque, NM (USA).
- [Thommesen and Andersen, 2012] Thommesen, J. and Andersen, H. B. (2012). *Human Error Probabilities (HEPs) for generic tasks and Performance Shaping Factors (PSFs) selected for railway operations DTU Management Engineering*. Technical edition.
- [Underwood and Waterson, 2013] Underwood, P. and Waterson, P. (2013). Systemic accident analysis: Examining the gap between research and practice. *Accident Analysis and Prevention*, 55:154–164, doi: <http://dx.doi.org/10.1016/j.aap.2013.02.041>.
- [Underwood and Waterson, 2014] Underwood, P. and Waterson, P. (2014). Systems thinking, the Swiss Cheese Model and accident analysis: A comparative systemic analysis of the Grayrigg train derailment using the ATSB, AcciMap and STAMP models. *Accident Analysis and Prevention*, 68:75–94, doi: <http://dx.doi.org/10.1016/j.aap.2013.07.027>.
- [UNISIG, 2012] UNISIG (2012). SUBSET-088-1 3.5.0 ETCS Application Level 1 - Safety Analysis: Part 2 - Functional Analysis. pages 1–52.
- [UNISIG (Union of Signalling Industry), 2016] UNISIG (Union of Signalling Industry) (2016). ETCS Application Level 2 - Safety Analysis Part 2 - Functional Analysis. SUBSET-088-2 Part 2. pages 1–61.
- [US NRC Regulation, 1975] US NRC Regulation (1975). Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants [NUREG-75/014 (WASH-1400)]. Technical report, US Nuclear Regulatory Commission Regulation. url: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr75-014/>.
- [U.S. Nuclear Regulatory Commission, 2002] U.S. Nuclear Regulatory Commission (2002). Review of Findings for Human Performance Contribution to Risk in Operating Events (NUREG/CR-6753, INEEL/EXT-01-01166). Technical report. url: <http://www.nrc.gov/reading-rm/doc-collections/nuregs/contract/cr6753/>.

-
- [Vanderhaegen, 1999a] Vanderhaegen, F. (1999a). APRECIH: a human unreliability analysis method – application to railway system. *Control Engineering Practice*, 7(11):1395–1403, doi: [http://dx.doi.org/10.1016/S0967-0661\(99\)00109-4](http://dx.doi.org/10.1016/S0967-0661(99)00109-4).
- [Vanderhaegen, 1999b] Vanderhaegen, F. (1999b). Cooperative system organisation and task allocation: illustration of task allocation in air traffic control. *Le Travail Humain*, 62(3):197–222.
- [Vanderhaegen, 2001] Vanderhaegen, F. (2001). A non-probabilistic prospective and retrospective human reliability analysis method — application to railway system. *Reliability Engineering & System Safety*, 71(1):1–13, doi: [http://dx.doi.org/10.1016/S0951-8320\(00\)00060-0](http://dx.doi.org/10.1016/S0951-8320(00)00060-0).
- [Vanderhaegen, 2010] Vanderhaegen, F. (2010). Human-error-based design of barriers and analysis of their uses. *Cognition, Technology & Work*, 12(2):133–142, doi: <http://dx.doi.org/10.1007/s10111-010-0146-3>.
- [Vanderhaegen, 2011] Vanderhaegen, F. (2011). Cooperation and learning to increase the autonomy of ADAS. *Cognition, Technology & Work*, 14(1):61–69, doi: <http://dx.doi.org/10.1007/s10111-011-0196-1>.
- [Vanderhaegen, 2014a] Vanderhaegen, F. (2014a). Dissonance engineering : a new challenge to analyse risky knowledge when using a system. *International Journal of Computers, Communications & Control*, 9(6):670–679.
- [Vanderhaegen, 2014b] Vanderhaegen, F. (2014b). Dissonance Engineering for Risk Analysis: A Theoretical Framework. In *Risk Management in Life-Critical Systems*, pages 157–181. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- [Vanderhaegen, 2016] Vanderhaegen, F. (2016). A rule-based support system for dissonance discovery and control applied to car driving. *Expert Systems with Applications*, 65:361–371, doi: <http://dx.doi.org/10.1016/j.eswa.2016.08.071>.
- [Vanderhaegen and Carsten, 2017] Vanderhaegen, F. and Carsten, O. (2017). Can dissonance engineering improve risk analysis of human–machine systems? *Cognition, Technology & Work*, 19(1):1–12, doi: <http://dx.doi.org/10.1007/s10111-017-0405-7>.
- [Vanderhaegen et al., 2010] Vanderhaegen, F., Cassani, M., and Cacciabue, P. (2010). Efficiency of safety barriers facing human errors. In *IFAC-HMS*, pages 1–6, Valenciennes, France.
- [Vanderhaegen and Caulier, 2011] Vanderhaegen, F. and Caulier, P. (2011). A multi-viewpoint system to support abductive reasoning. *Information Sciences*, 181(24):5349–5363, doi: <http://dx.doi.org/10.1016/j.ins.2011.07.050>.
- [Vanderhaegen and Zieba, 2014] Vanderhaegen, F. and Zieba, S. (2014). Reinforced learning systems based on merged and cumulative knowledge to predict human actions. *Information Sciences*, 276:146–159, doi: <http://dx.doi.org/10.1016/j.ins.2014.02.051>.
- [Vanderhaegen et al., 2011] Vanderhaegen, F., Zieba, S., Enjalbert, S., and Polet, P. (2011). A Benefit/Cost/Deficit (BCD) model for learning from human errors. *Reliability Engineering & System Safety*, 96(7):757–766, doi: <http://dx.doi.org/10.1016/j.res.2011.02.002>.
- [Villemeur, 1988] Villemeur, A. (1988). *Surete de fonctionnement des systemes industriels: fiabilite - facteurs humains, informatisation*. Dir. Et. Rech. Electr. France. Eyrolles, Paris.
- [Villemeur, 1992] Villemeur, A. (1992). *Reliability, Availability, Maintainability and Safety Assessment, Assessment, Hardware, Software and Human Factors*. Wiley, 2 edition.

-
- [Wang et al., 2011] Wang, A., Luo, Y., Tu, G., and Liu, P. (2011). Quantitative evaluation of human-reliability based on fuzzy-clonal selection. *IEEE Transactions on Reliability*, 60(3):517–527, doi: <http://dx.doi.org/10.1109/TR.2011.2161031>.
- [Weber et al., 2012] Weber, P., Medina-Oliva, G., Simon, C., and Iung, B. (2012). Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4):671–682, doi: <http://dx.doi.org/10.1016/j.engappai.2010.06.002>.
- [Whaley et al., 2011] Whaley, A. M., Kelly, D. L., Boring, R. L., and Galyean, W. J. (2011). SPAR-H Step-by-Step Guidance. Technical Report May, Idaho National Laboratory.
- [Wikipedia, 2017] Wikipedia (2017). European Train Control System — Wikipedia, the free encyclopedia. <http://en.wikipedia.org/w/index.php?title=European%20Train%20Control%20System&oldid=803929388>. [Online; accessed 30-October-2017].
- [Williams, 1985] Williams, J. (1985). HEART – A proposed method for achieving high reliability in process operation by means of human factors engineering technology. In *Proceedings of a Symposium on the Achievement of Reliability in Operating Plant*, Safety and Reliability Society (SaRS), NEC, Birmingham.
- [Williams, 1986] Williams, J. C. (1986). HEART—a proposed method for assessing and reducing human error. In *9th Advances in Reliability Technology Symposium*, University of Bradford. University of Bradford.
- [Wilson, 2014] Wilson, J. R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics*, 45(1):5–13, doi: <http://dx.doi.org/10.1016/j.apergo.2013.03.021>.
- [Wilson and Norris, 2005] Wilson, J. R. and Norris, B. J. (2005). Rail human factors: Past, present and future. *Applied ergonomics*, 36(6):649–60, doi: <http://dx.doi.org/10.1016/j.apergo.2005.07.001>.
- [Xu and Smets, 1996] Xu, H. and Smets, P. (1996). Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning*, 14(2-3):155–185, doi: [http://dx.doi.org/10.1016/0888-613X\(96\)00113-2](http://dx.doi.org/10.1016/0888-613X(96)00113-2).
- [Yager, 1987] Yager, R. R. (1987). On the dempster-shafer framework and new combination rules. *Information Sciences*, 41(2):93–137, doi: [http://dx.doi.org/10.1016/0020-0255\(87\)90007-7](http://dx.doi.org/10.1016/0020-0255(87)90007-7).
- [Zio, 2009] Zio, E. (2009). *Reliability engineering: Old problems and new challenges*, volume 94.
- [Zio and Ferrario, 2013] Zio, E. and Ferrario, E. (2013). A framework for the system-of-systems analysis of the risk for a safety-critical plant exposed to external events. *Reliability Engineering & System Safety*, 114(1):114–125, doi: <http://dx.doi.org/10.1016/j.ress.2013.01.005>.
-