# Learning from major accidents to improve system design

R. Moura*, M. Beer, E. Patelli & J. Lewis
*Institute for Risk and Uncertainty, University of Liverpool, United Kingdom*

F. Knoll
*NCK Inc., Montreal, Canada*

***Corresponding Author:***
R. Moura <rmoura@liverpool.ac.uk>
M: +44(0)7463 775979
*Office 614 Brodie Tower, Brownlow Street, Liverpool L69 3GQ, United Kingdom*

*Keywords:* accident analysis; MATA-D; human factors; human reliability analysis; risk & safety in design; CREAM

## 1 INTRODUCTION

### 1.1 *The human contribution to major accidents*

Recent major accidents in complex industrial systems, such as in oil & gas platforms and in the aviation industry, were deeply connected to human factors, leading to catastrophic consequences. A striking example would be the investigation report from the National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (2011) of the April 2010 blowout, in which eleven men died and almost five million barrels of oil were spilled in the Gulf of Mexico. The investigators unarguably emphasized the human factors role: features such a failure to interpret a pressure test and delay in reacting to signals were found to have interacted with poor communication, lack of training and management problems to produce this terrible disaster. Other contemporary investigation reports, such as the Rio-Paris Flight 447 (*Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile*, 2011) and Fukushima (Kurokawa, 2012), share the same characteristics regarding the significance of human-related features to the undesirable outcome.

Thus, the understanding of the interactions between human factors, technology aspects and the organisational context seems to be vital, in order to ensure the safety of engineering systems and minimise the possibility of major accidents. A suitable Human Reliability Analysis (HRA) technique is usually applied to approach the human contribution to undesirable events.

*National Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil.

### 1.2 *Human reliability analysis: a brief review*

Human Reliability Analysis (HRA) can be generally defined as a predictive tool, intended to estimate the probability of human errors and weigh the human factors contribution to the overall risk by using qualitative and/or quantitative methods.

In the early 60's, the first structured method to be used by industry to quantify human error was presented by Swain (1963), which later evolved to the well-known Technique for Human Error Rate Prediction - THERP (Swain and Guttman, 1983). This technique was initially developed to deal with nuclear plant applications, using in-built human error probabilities adjusted by performance-shaping factors and dependencies (interrelated errors) to deliver a human reliability analysis event tree. Some researchers (e.g. Reason, 1990; Kirwan, 1997; Everdij and Blom, 2013) refer to THERP as the most well-known method to assess human reliability and provide data to probabilistic safety assessments.

The accident model acknowledged as the "Swiss Cheese model", developed by Reason (1990), can be addressed as the most influential piece of work in the human factors field. It has been widely used to describe the dynamics of accident causation and explain how complex systems can fail through a combination of simultaneous factors (or as a result of the alignment of the holes of the Swiss cheese slices (Figure 1).

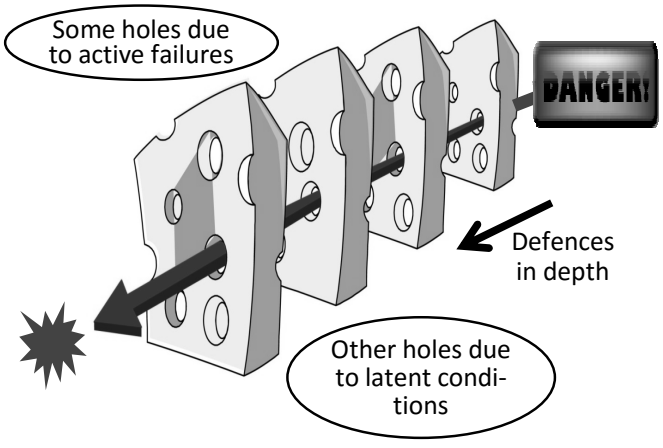


Figure 1. "Swiss Cheese Model" after Reason (1997)

Many Human Reliability Analysis subsequently developed were, to some extent, inspired by this model. Examples are the Human Factors Analysis Methodology – HFAM (Pennycook and Embrey, 1993), the Sequentially Outlining and Follow-up Integrated Analysis – SOFIA (Blajev, 2002), the Human Factors Analysis and Classification System – HFACS (Shappell et al. 2007), extensively used to investigate military and commercial aviation accidents, and the Systematic Occurrence Analysis Methodology - SOAM (Licu et al., 2007).

The concept that accidents arise from an arrangement of latent failures, later renamed to latent conditions (Reason, 1997), and active failures in complex systems demonstrated accuracy and practicality to guide prevention measures (Hopkins, 1999). Reason's studies of human errors have focused on the work environment, human control processes and safe operation of high-technology industrial systems, and included management issues and organisational factors.

There are several methods to assess human performance in different domains, and the development of such tools was notably triggered by the advances in high-technology industrial systems, particularly nuclear plants, aerospace, offshore oil and gas, military and commercial aviation, chemical and petrochemical, and navigation. Some of them were assessed by Bell and Holroyd (2009), who reported 72 different techniques to estimate human reliability and considered 35 to be potentially relevant. Further analysis highlighted 17 of these HRA tools to be of potential use for major hazard directorates in the United Kingdom. These techniques are usually separated by generations, which basically reflect the focus of the analysis.

The first generation methods, developed between the 60's and early 90's, are mainly focused on the task to be performed by operators. Essentially, potential human erroneous actions during the task sequence are identified, and the initial probability is then adjusted by internal and external factors (performance shaping factors, error-forcing conditions, scaling factors or performance influencing factors, depending on the methodology) to deliver a final estimation of human error probabilities. The key step in this approach is selecting the critical tasks to be performed by operators, which are considered to be elements or components subjected to failure due to inborn characteristics, thus having an "inbuilt probability of failure". These methods are widely recognised and commonly preferred by practitioners, probably because they provide a straightforward output such as an event tree or a probability value that can be directly integrated to Probabilistic Risk Assessments. Some examples are THERP, HEART (Human Error Assessment and Reduction Tech-

nique), presented by Williams (1986), and JHEDI (Justification of Human Error Data Information), introduced by Kirwan and James (1989).

Alternatively, second generation techniques have been developed from late 90's and are based on the principle that the central element of human factors assessments is actually the context in which the task is performed, reducing previous emphasis on the task characteristics *per se* and on a hypothetical inherent human error probability. "A Technique for Human Error Analysis" – ATHEANA (Cooper et al., 1996), the Connectionism Assessment of Human Reliability (CAHR) based on Sträter (2000) and the Cognitive Reliability and Error Analysis Method (CREAM) by Hollnagel (1998) are good examples of this kind of approach, all reflecting the focus shift from tasks to context to provide a better understanding of human error and integrate engineering, social sciences and psychology concepts. More recent literature (e.g. Kirwan et al., 2005; Bell and Holroyd, 2009) alludes to the Nuclear Action Reliability Assessment – NARA (Kirwan et al., 2005) as the beginning of the third generation methods. However, it seems to be merely an update of first generation techniques, i.e. HEART, using more recent data from newer databases such as CORE-DATA (Gibson and Megaw, 1999).

All these methods provide a number of taxonomies to handle possible internal and external factors that could influence human behaviour. Modern data classification taxonomies are mostly derived from Swain's (1982) work, in which he organised human errors in errors of omission and errors of commission, being the former a failure to execute something expected to be done (partially or entirely), while the latter can be translated as an incorrect action when executing a task or a failure to execute an action in time. The issue modelling human errors through the prediction of human behaviour during complex rare events was addressed by Rasmussen (1983), who envisioned the Skill-Rule-Knowledge (SRK) model. He differentiated three basic levels of human performance: skill-based, when automated actions follow an intention (sensory-motor behaviour); rule-based, when there is a procedure or technique guiding the action; and knowledge-based, represented by actions developed to deal with an unfamiliar situation. Reason (1990) split human errors in slips and lapses, when an execution failure or an omission occurs, and mistakes, which result from judgement processes used to select an objective, or the means to accomplish it. Later, Rasmussen's theory was encompassed by Reason to further categorise mistakes in rule-based mistakes, when a problem-solving sequence is known, but an error choosing the right solution to deal with the signals occurs; and knowledge-based mistakes, when the problem is not under a recognisable structure thus a stored troubleshooting solution cannot be immediately applied. Reason also highlighted an alternative behaviour from a social context, called "violation". This concept was split in exceptional and routine violations, both emerging from an intentional deviation from operating procedures, codes of practice or standards.

Although the classification schemes are usually connected to the industrial domain for which they were originally developed, some of them are non-specific (e.g. HEART) and thus have been successfully applied in a broader range of industries.

Regardless of the variety of HRA methods available to enable practitioners to assess the risks associated with human error by estimating its probability, the substantially high uncertainties related to the human behavioural characteristics, interlaced with actual technology aspects and organisational context, turn this kind of evaluation into a very complicated matter, thus has been raising reasonable concern about the accuracy and practicality of such probabilities.

### 1.3 *Human performance data limitations*

Data collection and the availability of a meaningful dataset to feed human reliability and other approaches related to the assessment of human performance in engineering systems seems to be the most severe constraints. Many studies in the early 90's addressed these issues, and both the unavailability of data on human performance in complex systems (Swain, 1990) and limitations related to the data collection process (International Atomic Energy Agency, 1990) were considered to be problems extremely difficult to overcome. Therefore, some efforts to collect accident data such as the Storybuilder Project (Bellamy et al, 2006) were undertaken, aiming at the classification and statistical analysis of occupational accidents.

In a contemporary review, Grabowski et al. (2009) suggests that the exponential rise of electronic records even worsened the problems related with human error data, stating that *data validation, compatibility, integration and harmonization are increasingly significant challenges in maritime data*

*analysis and risk assessments*. This indicates that difficulties to find usable human error and human factors data are still a major concern, which deserves to be carefully addressed by practitioners and researchers.

Moura et al (2015) discriminated some of the difficulties that might be preventing the development of a comprehensive dataset to serve as a suitable input to human performance studies in engineering systems. Main issues can be summarised by: (i) dissimilar jargons and nomenclatures used by distinct industrial sectors are absorbed by the classification method, making some taxonomies specific to particular industries; (ii) the effort to collect human data is time-consuming, and the need for inserting the "human contribution figures" into safety studies favours the immediate use of expert elicitation, instead of a dataset; (iii) The accuracy of the collection method is very difficult to assess, and distinct sources (e.g. field data, expert elicitation, performance indicators or accident investigation reports) can lead to different results; and (iv) the interfaces between human factors, technological aspects and the organisation are context-dependent and can be combined in numerous ways. This turns early predictions into a very difficult matter, due to the variability of the environment and the randomness of human behaviour.

Therefore, in this work, these significant drawbacks will be minimised by the development of a novel industrial accidents dataset, bringing together major accident reports from different industrial backgrounds and classifying them under a common framework. In spite of being a time-consuming and a laborious process, the accidents collection and the detailed interpretation will provide a rich data source, enabling the usage of integrated information to generate input to design improvement schemes.

Accident investigations can be considered to be one of the most valuable and reliable sources of information for future use, provided that several man-hours from a commissioned expert team are applied in an in-depth analysis of an undesirable event sequence, providing detailed insight into the genesis of industrial accidents.

## 2 CLASSIFICATION METHOD

### 2.1 *The Cognitive Reliability and Error Analysis Method (CREAM) taxonomy as a common framework to classify accidents*

In a previous work, some of the most used taxonomies in human reliability analysis were examined as possible inputs to the establishment of a data classification framework for a global accidents dataset. The three nomenclature sets considered by Moura (2015) were The Human Factors Analysis and Classification System - HFACS (Shappell et al. 2007), the Error Promoting Conditions (EPCs) from the Human Error Assessment and Reduction Technique (HEART) and the CREAM categorisation.

The fact that CREAM uses a nonspecific taxonomy, thus adaptable to most industrial segments, and its natural separation between man, technology and organisation, facilitating the accidents classification, made this terminology to be selected, in order to originate the structure of the new dataset.

Figures 2, 3 and 4 show the dataset classification structure.

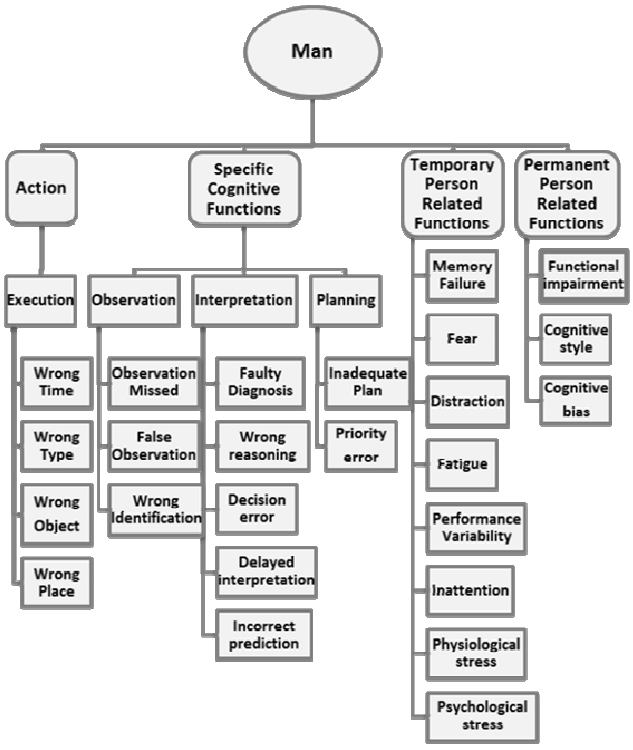


Figure 2. "Man" categorisation, adapted from Hollnagel (1998).

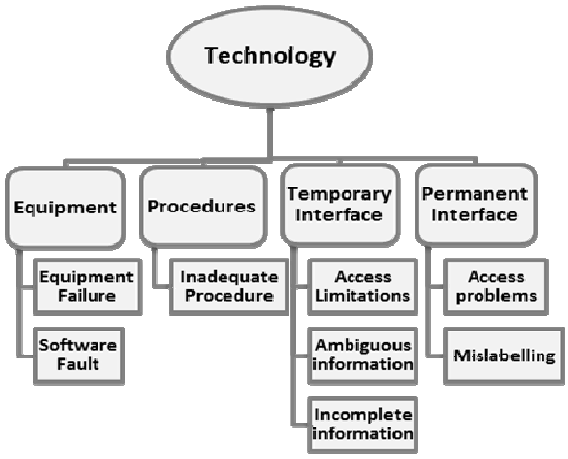


Figure 3. ""Technology" categorisation, adapted from Hollnagel (1998).

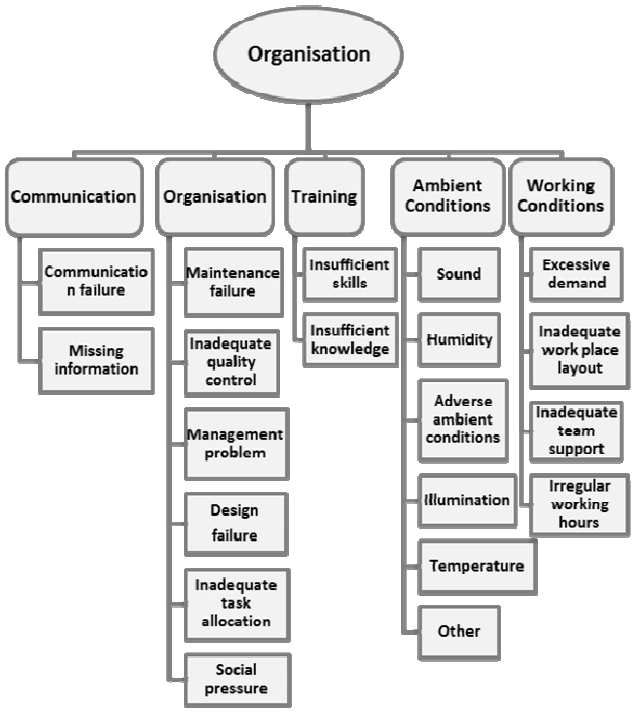


Figure 4. "Organisation" categorisation, adapted from Hollnagel (1998).

The 53 factors which could have influenced each of the 238 assessed accidents are organised in the three major groups depicted in Figures 2, 3 and 4. The "man" group concentrates human-related phenotypes in the action sub-group, representing the possible manifestation of human errors through erroneous actions (Wrong Time, Wrong Type, Wrong Object and Wrong Place), usually made by operators in the front-line. These flawed movements cover omitted or wrong actions; early, late, short, long or wrong movements, including in an incorrect direction or with inadequate force, speed or magnitude; skipping one or more actions or inverting the actions order during a sequence.

Possible causes or triggering factors with human roots can be classified as Specific Cognitive Functions, or the general sequence of mental mechanisms (Observe-Interpret-Plan) which leads the human being to respond to a stimulus. Also, temporary (e.g. fatigue, distraction or stress) and permanent disturbances (biases such as a hypothesis fixation or the tendency to search for a confirmation of previous assumptions) can be captured under the sub-groups Temporary and Permanent Person-related Functions. These are the person-related genotypes.

The second major group (Figure 3) represent technological genotypes, associated with procedures, equipment and system failures, as well as shortcomings involving the outputs (signals and information) provided by interfaces. The last group (Figure 4) encompasses organisational contributing factors, representing the work environment and the social context of the industrial activity. It involves latent conditions (such as a design failure), communication shortcomings and operation, maintenance, training, quality control and management problems. Factors such as adverse ambient conditions and unfavourable working conditions (e.g. irregular working hours) are also included in this category.

## 3 REVIEW OF 238 MAJOR ACCIDENTS: THE MULTI-ATTRIBUTE TECHNOLOGICAL ACCIDENTS DATASET (MATA-D)

### 3.1 *MATA-D conception: data selection*

To overcome the problems of the data collection process and the quality variability of different data sources, this work limits the data gathering to detailed accounts of accidents occurred in the industrial segments listed in Table 1. Accident reports and detailed case studies contain comprehensive information about the events, which can be interpreted and modelled into the groups and sub-groups shown in Figures 2, 3 and 4, to serve as input to the newly-created Multi-attribute Technological Accidents Dataset (MATA-D). The original reports were obtained from reliable sources such as regulators, investigation panels, government bodies, insurance companies and industry experts. A detailed account of the contributing institutions can be found in Moura et al (2015). The dataset covers major accidents occurred

worldwide, from the early fifties to today. Table 1 shows the accidents time-spam per industrial activity.

It is worth mentioning that the data selection criterion brought two significant gains. The use of real-life accounts reduces uncertainties related to the accurateness of the data, and investigation reports supply detailed technical info, evidences and an in-depth analysis of the interfaces between human factors, technology and the organisation in which the event occurred. This seems to be one of the finest sources of information available, from which MATA-D is fully designed.

Table 1. MATA-D events distribution by industry

| Industry | Accidents | | Period |
|---|---|---|---|
| | # | % | |
| Refinery | 39 | 16.39 | 1978 - 2011 |
| Upstream (Oil & Gas) | 37 | 15.55 | 1975 - 2012 |
| Chemicals Factory | 29 | 12.18 | 1975 - 2011 |
| Petrochemicals | 25 | 10.50 | 1974 - 2008 |
| Nuclear Industry | 23 | 9.66 | 1953 - 2011 |
| Civil Construction | 16 | 6.72 | 1968 - 2011 |
| Terminals and Distribution | 15 | 6.30 | 1975 - 2012 |
| Aviation Industry | 13 | 5.46 | 1996 - 2013 |
| Gas Processing | 09 | 3.78 | 1977 - 2008 |
| Metallurgical Industry | 07 | 2.94 | 1975 - 2011 |
| Waste Treatment Plant | 05 | 2.10 | 2002 - 2009 |
| Food Industry | 04 | 1.68 | 1998 - 2009 |
| Other | 16 | 6.72 | 1980 - 2011 |

### 3.2 *MATA-D usage*

This new accident dataset aims to provide researchers and practitioners with a simple and innovative interface for classifying accidents from any industrial sector, reflecting apparently dissimilar events in a comparable fashion. The binary classification for the evaluated factors (presence or absence) allows data interpretation using uncomplicated statistical methods or sophisticated mathematical models, depending on the user's requirements.

Moreover, the detailed descriptions available for each identified factor, as can be seen in the example given in Table 2, allows comprehensive understanding and analysis of single accidents, as well as the disclosure of the precise evidence of failures associated with psychological (cognitive functions), engineering (e.g. design and equipment failures) and organisational (e.g. management problems and training) aspects. These descriptions provide an effective translation of highly technical content reports to a linguistic approach easily understood by practitioners from outside the engineering field, facilitating cross-disciplinary communication among professionals and academics. Many applications can be developed from these unique characteristics.

### 3.3 *Features of the data sample*

1,539 fatalities were recorded in 67 of the 238 analysed events. Some of the reports also contained damage recovery information, and 95 events were accountable for more than £20 billion in material losses. Apart from these significant features, it is acknowledged that many additional costs arise from major accidents. It is reported (Fowler, 2013) that British Petroleum (BP) paid around US$ 14 billion in indenisations related to the Gulf of Mexico oil spill clean-up, and Bell (2012) described a 35.00% stock price drop from the event occurrence in 2010 to 2012.

However, the most significant feature of the dataset events is that all of them involved a major emission, fire, explosion or crash, exposing humans and/or the environment to serious danger. Thus, these events largely fit the definition of "major accident", according to the United Kingdom's Control of Major Accident Hazards Regulations (1999).

### 3.4 *MATA-D Construction: data interpretation and classification demonstration method*

The analysis and classification of nearly 250 accident reports (some events were investigated by more than one entity and had multiple records) was a time-consuming process, but enabled the comparison among accidents from different industries. Investigation reports varied from a few to a maximum of 494 pages.

The process involved the interpretation of the accidents reports and their subsequent classification under the common taxonomy to create the dataset.

Table 2 exemplifies how one of the collected accidents was carefully decomposed and recorded in the MATA-D database. This example scrutinises a severe explosion of flammable gasoline constituents released from a refinery's hydrofluoric acid (HF) alkylation unit, examined by the US Chemical Safety and Hazard Investigation Board (2005). The release, ignition, fire and several explosions occurred during the preparation of a pump repair, which was being removed by maintenance workers. As a consequence, six employees were injured, the production was stopped for approximately 6 months and a damage repair cost of US$ 13 million was reported.

Table 2. Oil Refinery Fire and Explosion classification example

| Group | Sub-Group | Factor | Description* |
|---|---|---|---|
| Man | Execution | Wrong Type | Movement in the wrong direction: during a seal repair, the operator attempted to isolate the pump by closing a plug valve. He moved the valve wrench to a perpendicular position in relation to the flow, believing this was the closed position, but the valve was actually open. |
| Man | Specific Cognitive Functions (Observation) | Observation missed | Overlook cue/signal: The valve stem was equipped with a position indicator, but the operators overlooked it. The indicator was correctly indicating the open status. |
| Man | Specific Cognitive Functions (Observation) | Wrong Identification | Incorrect identification: the mechanic specialist recognised the valve as closed due to the wrench position and, following a safety procedure, placed locks and tags on the valve, to prevent its inadvertent opening. |
| Man | Specific Cognitive Functions (Interpretation) | Wrong reasoning | (i) Deduction error: Operator and mechanic specialist firmly believed that the closed valve position was always identified by the wrench being perpendicular to the flow of product.<br>(ii) Induction Error: after unbolting the flare line, a small release of a high flammable component was observed for a few seconds. As the flux stopped, the operator inferred that the pump was de-pressurised and the removal was safe. However, vent line was clogged by scale. |
| Man | Specific Cognitive Functions (Planning) | Priority error | After the installation of the locks, the operator noticed that the position indicator was showing that the valve was open, but he maintained his plans and left the plant to fetch the necessary tools for the pump removal. |
| Man | Permanent Person Related Functions | Cognitive bias | Confirmation bias: search for information was restricted to looking at wrench position, which confirmed the operator's assumption that the valve was closed, dismissing a further consideration of the fully functioning position indicator. |
| Technology | Temporary Interface | Ambiguous Information | Position mismatch: wrench collar was installed in an unusual position. Usually, the perpendicular wrench position indicates the closed state, while the parallel wrench position indicates the open status. Thus, wrench position (open-close) was inverted and thus conflicting with the position indicator. |
| Organisation | Organisation | Maintenance Failure | (i) There was no effective preventive/predictive maintenance programme to maintain pumps operational, as interventions (repair / parts replacement) took place only when equipment failed. The investigation of possible failure mechanisms (the actual causes of the breakdowns) never occurred.<br>(ii) Flare line was clogged by scale. |
| Organisation | Organisation | Inadequate Quality Control | (i) Despite the recurring failures of several pump seals in the plant (prior to the accident, 23 work orders for similar defects were issued), quality control procedures failed to ensure the adequacy of the equipment to the transported product and to certify that maintenance procedures were suitable.<br>(ii) Quality control failed to identify the inadequate installation of the wrench collar, which allowed the wrench bar to stay in an unusual position, unfamiliar to operators. |
| Organisation | Organisation | Design Failure | The valve actuator (wrench) collar had a squared shape and could be installed in any position, thus there was a discrepancy between the design of the valve and its actuator. Design should have prevented the wrench installation in an unusual position. Also, further investigation identified that the original actuator was a gear-operated one, and the design change to a wrench actuator failed to address further safety implications, such as the produced mismatch between the position indicator and the wrench position. |

* Adapted from the evidence/accounts from the US Chemical Safety and Hazard Investigation Board (2005) Case Study.

The classification method was applied to the above accident and the Table 2 clearly exemplifies how the investigation report from an event occurred in a specific industry (i.e. a refinery) can be decomposed into general categories, enabling the association with most industrial sectors. This classification method allowed the creation of a dataset composed by major accidents from industries with no apparent connection, but sharing common features (groups, subgroups and factors) which contributed to serious events. In addition, the dataset preserves the main characteristics of the scrutinised events at the description column, facilitating the prompt understanding of complex investigation reports and allowing further analysis of single or grouped events, if required.

A seemingly pure human error (which could be described by the removal of the pump without closing the isolation valve or, more specifically, opening the isolation valve instead of closing it) can be explained by some cognitive mechanisms triggered by technology and organisational issues. The worker tried to isolate a pump for maintenance by putting the valve wrench in a perpendicular position in relation to the piping, which is a widely accepted convention for the closed state of a valve. He disregarded the position indicator at the valve body, assuming the wrench position as a sufficient proof of the pump isolation. A mechanic specialist who was responsible for double-checking the isolation, for safety reasons, also deduced that the valve was closed just by looking at the wrench position, and locked the valve. This allows the identification of valuable cognitive functions influencing the human erroneous actions, assisted by the terminology of the classification method, such as the observation missed, the wrong identification, wrong reasoning and priority error. A person-related cognitive bias was also categorised, explaining why the operator ignored the position indicator.

Even more important, the link between technology, design and human factors can be clearly established: the ambiguity of the information provided by the interface (unfamiliar wrench position versus position indicator), triggered by the design failure, motivated the operator to reason in a way that the error of opening the isolation valve, instead of closing it, was plausible. Other organisational contributors, such as the quality control faults of the wrench installation and the mechanical integrity programme, were also captured by the classification scheme.

### 3.5 *MATA-D Results & Analysis*

Following the same method presented in Table 2, 238 major accidents were scrutinised and computed into the MATA-D. Tables 3, 4, 5 and 6 summarise the results obtained from the interpretations of the accident reports analysed by the authors, as well as the resulting categorisation.

Table 3. Data Classification results (main groups).

| Group | Frequency* | |
|---|---|---|
| | # | % |
| Man | 136 | 57.14 |
| Technology | 196 | 82.35 |
| Organisation | 227 | 95.38 |

*Number of events where groups appeared.

Table 4. Data Classification results (factors & subgroups).

| Factor | Frequency* | | Sub-Group | Freq.* |
|---|---|---|---|---|
| | # | % | | % |
| Wrong Time | 35 | 14.70 | Execution | 54.60 |
| Wrong Type | 28 | 11.80 | | |
| Wrong Object | 06 | 2.50 | | |
| Wrong Place | 75 | 31.50 | | |
| Observation Missed | 37 | 15.50 | Cognitive | |
| False Observation | 08 | 3.40 | Functions** | 47.50 |
| Wrong Identification | 06 | 2.50 | | |
| Faulty diagnosis | 31 | 13.00 | | |
| Wrong reasoning | 27 | 11.30 | | |
| Decision error | 22 | 9.20 | | |
| Delayed interpretation | 11 | 4.60 | | |
| Incorrect prediction | 09 | 3.80 | | |
| Inadequate plan | 23 | 9.70 | | |
| Priority error | 17 | 7.10 | | |
| Memory failure | 02 | 0.90 | Temp Person | |
| Fear | 05 | 2.10 | Related | |
| Distraction | 14 | 5.90 | Functions | 13.00 |
| Fatigue | 07 | 2.90 | | |
| Performance Variability | 03 | 1.40 | | |
| Inattention | 05 | 2.10 | | |
| Physiological stress | 02 | 0.80 | | |
| Psychological stress | 07 | 2.90 | | |
| Functional impairment | 01 | 0.40 | Perm. Person | |
| Cognitive style | 00 | 0.00 | Related | |
| Cognitive bias | 17 | 7.10 | Functions | 7.60 |
| Equipment failure | 131 | 55.00 | Equipment | 56.30 |
| Software fault | 06 | 2.50 | | |
| Inadequate procedure | 105 | 44.10 | Procedures | 44.10 |
| Access limitations | 03 | 1.30 | Temporary | |
| Ambiguous information | 06 | 2.50 | Interface | 18.90 |
| Incomplete information | 42 | 17.60 | | |
| Access problems | 04 | 1.70 | Permanent | |
| Mislabelling | 04 | 1.70 | Interface | 3.40 |
| Communication failure | 25 | 10 .50 | Communi- | |
| Missing information | 49 | 20.60 | cation | 29.00 |

| | | | | |
|---|---|---|---|---|
| Maintenance failure | 83 | 34.90 | Organisation | 94.10 |
| Inadequate quality control | 144 | 60.50 | | |
| Management problem | 22 | 9.20 | | |
| Design failure | 157 | 66.00 | | |
| Inadequate task allocation | 143 | 60.10 | | |
| Social pressure | 17 | 7.10 | | |
| Insufficient skills | 86 | 36.10 | Training | 54.20 |
| Insufficient knowledge | 84 | 35.30 | | |
| Temperature | 03 | 1.30 | Ambient | |
| Sound | 00 | 0.00 | Conditions | 8.80 |
| Humidity | 00 | 0.00 | | |
| Illumination | 02 | 0.80 | | |
| Other | 00 | 0.00 | | |
| Adverse ambient condition | 17 | 7.10 | | |
| Excessive demand | 13 | 5.50 | Working | |
| Poor work place layout | 06 | 2.50 | Conditions | 11.30 |
| Inadequate team support | 08 | 3.40 | | |
| Irregular working hours | 09 | 3.80 | | |

*Number of events where factors or sub-groups appeared.
** Cognitive functions detailed on Table 5.

Table 5. Data Classification results (cognitive functions).

| Cognitive Function | Frequency* | |
|---|---|---|
| | # | % |
| Observation | 47 | 19.70 |
| Interpretation | 79 | 33.20 |
| Planning | 38 | 16.00 |

*Number of events where cognitive functions appeared.

Tables 3, 4 and 5 specify the number of appearances of the man-related, technology and organisational phenotypes and genotypes identified in the major accidents examined. Percentages relate to the total of events (238).

At least one human element was identified in 57.14% of the cases, with 54.60% of direct erroneous actions (phenotypes). Cognitive functions accounted for 47.50%, with the interpretation genotype appearing as the most relevant (33.20%). At least one technology genotype was recognised in 82.35% of the accidents, highlighting equipment failure (55.00%) and inadequate procedures (44.10%) as the foremost factors related to this group. Organisational issues appeared in 95.38% of the accidents, emphasising design failures (66.00%), inadequate quality control (60.50%) and inadequate task allocation (60.10%) as the most significant genotypes within the group.

Table 6 presents a macro-analysis of the major groups (man, machine and organisation), indicating that a single group causing a major accident is not common. Merely 0.84% of the examined events showed an erroneous action with a man-related genotype resulting in an accident. Exclusively technological factors were responsible for the undesirable outcome in only 3.78% of the cases, while 7.56% of the events were solely explained by organisational factors. On the other hand, combinations involving a minimum of two groups featured significantly in the dataset. A Man-Technology arrangement appeared in 47.48% of the cases, while a Man-Organisation combination performed in 56.30%. The Technology-Organisation pair figured together in 78.57% of the events. In 47.48% of the cases, the three groups appeared together. Table 6 summarises these results.

Table 6. Macro-analysis (main groups).

| Group / Combination | Frequency* | |
|---|---|---|
| | # | % |
| Only Man | 02 | 0.84 |
| Only Technology | 09 | 3.78 |
| Only Organisation | 18 | 7.56 |
| Man-Technology | 113 | 47.48 |
| Man-Organisation | 134 | 56.30 |
| Technology-Organisation | 187 | 78.57 |
| Man-Technology-Organisation | 113 | 47.48 |

*Events where a single group or combinations appeared.

There is a close relationship between the design failure genotype and the man group: 72.80% of the erroneous actions (execution errors) were accompanied by a design failure, such as in the case study presented on Table 2. In addition, 62.50% of temporary and permanent person related functions and 74.34% of cognitive functions were connected to design failures.

Also, it is important to notice that the design failure is the most significant single genotype from all three groups, appearing with an incidence of 66.00%, followed by inadequate quality control (60.50%), inadequate task allocation (60.10%) and equipment failure (55.00%). Despite the significance of these further contributing factors, which can be used in future studies to improve the organisation of work and disclose operational strategies, the following discussion will focus on the design failure genotype features and connections revealed by the statistical analysis.

## 4 DISCUSSION

### 4.1 *Improving robustness of system design*

Design failures were detected in 157 of the 238 major accidents included in the MATA-D, clearly emphasising the need for further developments in design verification schemes. These deficiencies are examples of embedded failures in the system design, which can stay dormant for many years before aligning with human errors, technology issues and other organisational problems to result in a serious occur-

rence. The failures related with the design of the Fukushima nuclear power plant, such as insufficient tsunami defences combined with the lack of flood protection for batteries, which caused the loss of DC power, remained dormant for decades. Similarly, icing problems of the original speed sensors in the Airbus 330 airplane persisted for approximately 8 years before triggering the catastrophic Rio-Paris flight crash in 2009. Although these design flaw examples could be promptly addressed (before the alignment of the holes in the Figure 1), the lack of a robust dataset containing useful information about the multifaceted interaction between human factors, technology and organisation in complex systems may be preventing standards and regulations from addressing the human performance problem in earlier stages of the lifecycle of engineering systems, such as design, in a structured way. The MATA-D construction intends to break this tendency, being composed by major accidents from high-technology industries to create means of analysing this kind of catastrophic events. Also, major accidents are notably rare events, and the wide-ranging taxonomy used to classify events in the MATA-D allows the accumulation of data from several industrial sectors to perform a deeper analysis and disclose early contributors and significant tendencies leading to human errors.

Other studies were able to identify this relationship between human errors, technology and organisational issues. Bellamy et al (2013) analysed 118 incidents involving loss of containment in Dutch Seveso plants and identified that 59% of the failures to use/operate a safety barrier were associated with human errors. Despite the application of a different classification system for human errors and the inclusion of events with minor consequences (only 9 out of 118 events were major accidents reportable under Seveso II Directive), these figures might well be related with the present study findings, in which 57.14% of the 238 major industrial accidents were found to have human contributing causes, as reflected in the "man" category statistics.

The comparison of single group accidents with the statistics for at least two simultaneous groups on Table 6 confirms that high-technology systems require a complex interaction of multiple failures in order to produce a major accident. It is important to notice that not only a number of barriers need to be breached, but it also has to interact in a very particular way. This makes the prediction of all design interactions and responses to human, technology and organisational events virtually impossible, highlighting the importance of developing design verification schemes to raise the awareness level of designers relating to major accidents. Therefore, providing some straightforward information based on the most common interactions occurring in complex accidents may be of assistance. The relationship between design failures and human factors indicates that the design damage tolerance criteria must be tested against specific human-related factors disclosed by this research. The direct association of execution errors and cognitive functions with design problems is a valuable finding, demonstrating how design failures can deeply influence human behaviour.

Design failures particularly appear to trigger failures in the human capacity to interpret system status (wrong reasoning and faulty diagnosis), enable potential observation misses and cause some execution errors (sequence, timing and type).

Based on these findings, an effective design review process should carefully address circumstances where some system analysis/diagnosis, interpretation or hypothesis formulations are required before taking an action. The common man, technology and organisation interfaces discussed indicate that it is likely that cues, measurements or information originally intended to lead to a human action have a substantial probability of being missed, an effect explained by some specific cognitive functions (inferences, generalisations or deductions) highlighted by this study.

The aim of the review would be to improve system design by making it responsive to common active failures translated as human erroneous actions, such as omissions; jumping forward a required action; performing a premature, delayed or wrong action; and performing a movement in the wrong direction, with inadequate speed or magnitude. Of course these operators' "action failures" occur in a greater frequency than accidents, and should be considered customary, or part of a non-mechanic behaviour. Consequently, human performance will vary, and it seems that addressing design shortcomings which can affect human behaviour, by learning from major accidents in an informed and structured way, is a reasonable path to reduce major accidents and tackle the genesis of human errors.

### 4.2 *Using the MATA-D for a design review process: an example*

One suitable example of effective design improvement approach would be to apply a design review process which considers the connections between

human erroneous actions, cognitive functions and design failures highlighted during this study.

The role of the proposed review process is to identify and correct design imperfections that could lead to major accidents. Primarily, due to the complexities of high-technology systems, it is important to bear in mind that one reviewer is unlikely to hold all necessary knowledge to assess all design disciplines and aspects (system functionalities, materials, mechanics, structure, fabrication methods, electrics, chemistry, corrosion protection, risk, compliance etc). The person in charge should be able to form a team, identifying and engaging with experts in the respective fields (face-to-face meetings), whether or not they are directly involved in the business. Designers, manufacturers, constructors and operators, for instance, are obvious interested parties, but referring to external parties, such as associations, academic institutions and regulatory bodies, will also aggregate significant value to the group task, being "time" the key constraint to be managed during this phase.

In summary, the first step would be to *(i)* identify and rank the safety critical elements (SCE) within the installation. One helpful definition of SCEs is found in the UK Safety Case Regulations (2005), in which the term is defined as any part of an installation whose failure could cause or contribute to a major accident, or elements designed to prevent or limit the effect of a major accident. Considering the wide range of high-technology installations encompassed by the MATA-D (e.g. oil and gas, nuclear plants and aviation), the SCEs list will vary enormously from facility to facility, depending on the industrial segment assessed. Then, *(ii)* the information associated with the critical elements (e.g. material and functionalities description, conceptual and detailed design, fabrication and installation drawings and process and instrumentation diagrams) are used to disclose the relevant human tasks, and *(iii)* the identified operations would be tested against the basic execution errors disclosed by this study (i.e. omissions; jumping forward a required action; performing a premature, delayed or wrong action; and performing a movement in the wrong direction, with inadequate speed or magnitude), to identify undesirable effects affecting the critical elements documented in step *(i)*. Next stage would involve *(iv)* the assessment of indications intended to trigger human actions, such as cues, measurements and displays. The possibility of missing them, as discussed in previous sections, should prompt deep consideration about the alternative measures in place (e.g. redundancy, double-check, automatic shut-down, supervisor intervention) to provoke human responsiveness. The last review step would comprise the *(v)* analysis of complex tasks, which can be defined as the ones requiring observation of signals, its correct interpretation and system diagnosis.

The mental modelling is inherent to the worker's level of knowledge, the information available and the work environment/situation, among other factors, thus the matter of a human inadequate reasoning while evaluating relevant conditions linked to critical elements must be considered in the review. Although this may seem, at first glance, an excessively challenging task to be undertaken by the design reviewer, the MATA-D results, which indicate specific mechanisms leading to poor interpretations, can be used to build a systematic assessment process. The available inputs to diagnose the undesirable condition should be listed and evaluated, in order to identify where: a) information (e.g. instructions, codes & standards, manuals, signals, communication); b) knowledge (e.g. level of training, education and engineering practice) and c) the work situation (e.g. adverse ambient conditions, irregular working hours, and inadequate work place layout) are likely to induce inferences, generalisations or deductions which can lead to invalid results.

Also, most of the industrial fields allow the designers to choose among a wide range of standards and protocols as an input to design. Thus, compliance verification is similarly a significant method to detect information imperfections, i.e. if the engineering best practices for the existing condition are being applied. The usage of codes and standards which consider human factors as well as the disclosed interactions should be preferred.

Consequently, this design review process should be able to identify possible blind spots and reflect a "design clarity" degree, indicating if the expected functions defined in the conceptual phase are thoroughly satisfied during the earlier stages of the installation lifecycle.

## 5 CONCLUSIONS

### 5.1 *A new method to apply past accidents lessons to design reviews*

Learning from past accidents is essential to minimise the possibility of undesirable events recurring, but this is not a trivial task. The particular sequence of events resulting in a serious accident is multidimensional and highly associated with the perfect align-

ment of very specific circumstances within a work environment. Consequently, limited learning is likely to arise from the analysis of a single event or even a few accidents, justifying the need for a broad comprehensive understanding of the common features and mechanisms leading to human error, which is the aim of the large major accidents collection.

A new accident dataset, created from detailed investigation reports and using a classification that admits events from different industries, was then introduced.

This work also described some advantageous findings for designers and practitioners who deal with major hazard control, in the sense that it is essential to take human error into account during design. Accordingly, improved insight into erroneous actions and influencing factors was revealed, as the vast collection of real-life accidents (i.e. 238) presented relevant relationships between man, technology and organisation and disclosed common patterns within disasters from different industrial segments.

Specific human factors to be addressed in a design review were then presented in the discussion section guidelines, reducing the burden and the time required to apply extensive human error lists to predicted tasks or complicated methodologies during the development of new projects. This approach, due to its simplicity, can be easily adapted to current design review processes, effectively raising awareness for the development of strategies to minimise human error through design.

The MATA-D includes valuable lessons from several high-technology industries, such as upstream, refining, aviation and nuclear, involving specialists from different fields and providing common input to major hazard control strategies. This new dataset can be used for any application requiring technical input from past major accidents.

## 6 ACKNOWLEDGEMENTS

This study was partially funded by CAPES (Proc. nº 5959/13-6).

## 7 REFERENCES

Bell, J. 2012. The Gulf Spill: BP Still Doesn't Get It. In Allen, F. E. (ed), *Forbes, 20 April 2012*. http://www.forbes.com/sites/frederickallen/2012/04/20/the-gulf-spill-bp-still-doesnt-get-it/

Bell J & Holroyd J. 2009. *Review of human reliability assessment methods*. Suffolk: HSE Books.

Bellamy L.J. et al. 2007. Storybuilder — A tool for the analysis of accident reports. *Reliability Engineering and System Safety* 92: 735-744.

Bellamy, L.J. et al. 2013. Analysis of underlying causes of investigated loss of containment incidents in Dutch Seveso plants using the Storybuilder method. *Journal of Loss Prevention in the Process Industries* 26: 1039-1059.

Blajev, T. 2002. *SOFIA (Sequentially Outlining and Follow-up Integrated Analysis) Reference Manual*. Brussels: EATMP Infocentre.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. 2011. *Final Report on the accident on 1st June 2009 to the Airbus A330-203* [Online]. Available from: http://www.bea.aero/docspa/2009/f-cp090601 .en/pdf/f-cp090601.en.pdf (Accessed: 06 November 2014).

Cooper, S. et al. 1996. *A Technique for Human Error Analysis (ATHEANA) - Technical Basis and Methodology Description.* Washington, D.C.: US Nuclear Regulatory Commission Library.

Everdij, M. & Blom, H. 2013. *Safety Methods Database version 1.0* [Online]. Amsterdam: National Aerospace Laboratory (NLR). Available from: http://www.nlr.nl/downloads/safety-methods-database.pdf (Accessed: 9 April 2014).

Fowler, T. 2013. BP Faces New Bout of Spill Liability. *The Wall Street Journal, 18 February 2013*. New York: Dow Jones & Company, Inc.

Gibson W. H. & Megaw T. D. 1999. *The implementation of CORE-DATA, a computerised human error probability database.* Suffolk: HSE Books.

Grabowski, M. et al. 2009. Human and organizational error data challenges in complex, large-scale systems. *Safety Science* 47: 1185-1194.

Hollnagel, E. 1998. *Cognitive Reliability and Error Analysis Method*. Oxford: Elsevier Science Ltd.

Hopkins, A. 1999. The limits of normal accident theory, *Safety Science*, 32, pp. 93-102.

International Atomic Energy Agency. 1990. Human Error Classification and Data Collection. *Report of a technical committee meeting organised by the IAEA, Vienna, 20-24 February 1989*. Vienna: INIS Clearinghouse.

Kirwan, B. 1997. Validation of human reliability assessment techniques: Part 1 — Validation issues, *Safety Science*, 27, pp. 25-11.

Kirwan, B. and James, N. J. 1989. Development of a human reliability assessment system for the management of human error in complex systems. *Proceedings of the Reliability '89, Brighton, 14-16 June*: pp. 5Al2/1-5A/2/11.

Kirwan, B. et al. 2005. Nuclear action reliability assessment (NARA): a data-based HRA tool, *Safety & Reliability*, 25, No. 2, pp. 38–45.

Kurokawa, K. et al. 2012. *The Official Report of The Fukushima Nuclear Accident Independent Investigation Commission - Executive Summary* [Online] Tokyo: The National Diet of Japan. Available from: https://www.nirs.org/fukushima/naiic_report.pdf (Accessed: 6 November 2014).

Licu, T. et al. 2007. Systemic Occurrence Analysis Methodology (SOAM) - A "Reason"-based organisational methodology for analysing incidents and accidents, *Reliability Engineering and System Safety*, 92, pp. 1162-1169.

Moura, R. et al. 2015. Human error analysis: Review of past accidents and implications for improving robustness of system design, Nowakowski, T. et al. (Eds), *Proceedings of the 24th European Safety and Reliability Conference, 14-18 September 2014, Wroclaw*. London: Taylor & Francis Group, pp. 1037-1046.

National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling. 2011. *The Gulf Oil Disaster and the Future of Offshore Drilling – Report to the President* [Online] Washington D.C.: U.S. Government Printing Office. Available from: http://www.gpo.gov/fdsys/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf (Accessed: 20 July 2015).

Pennycook, W. & Embrey, D. 1993. 'An operating approach to error analysis', *Proceedings of the First Biennial Canadian Conference on Process Safety and Loss Management*, April, Edmonton, Canada.

Rasmussen, J. 1983. Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models, *IEEE Transactions on Systems, Man and Cybernetics* 3, May, vol. SMC-13.

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

Reason, J. 1997. *Managing the Risks of Organizational Accidents.* Brookfield, USA: Ashgate.

Shappell, S., et al. 2007. Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System. *Human Factors* 49(2): 227-242.

Sträter, O. 2000. *Evaluation of Human Reliability on the Basis of Operational Experience*. Cologne: GRS. (English translation of the Report GRS-138: Beurteilung der menschlichen Zuverlassigkeit auf Basis von Betriebserfahrung.)

Swain, A. 1963. *A Method for Performing Human Factors Reliability Analysis,* Monograph-6851, Albuquerque: Sandia National Laboratories.

Swain, A. 1982. Modeling of Response to Nuclear Power Plant Transients for Probabilistic Risk Assessment, *Proceedings of the 8th Congress of the International Ergonomics Association*, August, Tokyo.

Swain, A. 1990. Human Reliability Analysis - Need, Status, Trends and Limitations. *Reliability Engineering and System Safety* 29: 301-313.

Swain, A. & Guttmann, H. 1983. *NUREG/CR 1278 - Handbook of Human Reliability Analysis with Enphasis on Nuclear Power Plant Applications*. Albuquerque: Sandia National Laboratories.

*The Control of Major Accident Hazards Regulations*. 1999. http://www.legislation.gov.uk/uksi/1999/743/contents/made. Surrey: National Archives.

*The Offshore Installations (Safety Case) Regulations*. 2005. http://www.legislation.gov.uk/uksi/2005/3117/made. Surrey: National Archives.

US Chemical Safety and Hazard Investigation Board 2005. *Case Study 2004-08-I-NM Oil Refinery Fire and Explosion*. Washington, DC: CSB Publications.

Williams, J.C. 1986. HEART - A Proposed Method for Assessing and Reducing Human Error. *Proceedings of the 9th Advances in Reliability Technology Symposium, Bradford, 2-4 April 1986.* Warrington: National Centre of Systems Reliability.