

Tracing New Safety Thinking Practices in Safety Investigation Reports

Nektarios Karanikas^{1,*} and Dimitrios Chionis²

¹ Aviation Academy, Amsterdam University of Applied Sciences, the Netherlands

² Psychology Department, Bolton University at New York College, Greece

ABSTRACT

Modern safety thinking and models focus more on systemic factors rather than simple cause-effect attributions of unfavourable events on the behaviour of individual system actors. This study concludes previous research during which we had traced practices of new safety thinking practices (NSTPs) in aviation investigation reports by using an analysis framework that includes nine relevant approaches and three safety model types mentioned in the literature. In this paper, we present the application of the framework to 277 aviation reports which were published between 1999 and 2016 and were randomly selected from the online repositories of five aviation authorities. The results suggested that all NSTPs were traceable across the sample, thus followed by investigators, but at different extents. We also observed a very low degree of using systemic accident models. Statistical tests revealed differences amongst the five investigation authorities in half of the analysis framework items and no significant variation of frequencies over time apart from the Safety-II aspect. Although the findings of this study cannot be generalised due to the non-representative sample used, it can be assumed that the so-called new safety thinking has been already attempted since decades and that recent efforts to communicate and foster the corresponding aspects through research and educational means have not yet yielded the expected impact. The framework used in this study can be applied to any industry sector by using larger samples as a means to investigate attitudes of investigators towards safety thinking practices and respective reasons regardless of any labelling of the former as “old” and “new”. Although NSTPs are in the direction of enabling fairer and more in-depth analyses, when considering the inevitable constraints of investigations, it is more important to understand the perceived strengths and weaknesses of each approach from the viewpoint of practitioners rather than demonstrating a judgmental approach in favour or not of any investigation practice.

Keywords: Safety Thinking; Safety Investigations; Safety Models; Accident Models.

1. INTRODUCTION

Modern system complexity emerging from the multiple interactions amongst technology, human agents, and organisational aspects (Martinetti et al., 2018) has driven safety thinking advancements with a focus more on systemic factors rather than components. Safety perspectives that interpret adverse events merely as results of human errors are linked with tendencies to (in)directly blame underperforming individuals, evaluate system performance levels based on a small number of unfortunate events, and neglect the daily successes of safe practices at the work floor under the reality of conflicting goals or varying conditions. This set of views has been described by Hollnagel (2013) as ‘Safety I’ and by Dekker (2007) as ‘Old View’ and is most frequently linked to safety investigations. On the other hand, the new safety thinking advocates a more systemic and human-centric approach to safety with the goal to understand better how

* Corresponding author: +31621156287, n.karanikas@hva.nl, nektkar@gmail.com

socio-technical systems function to achieve their objectives and how we could foster their strengths instead of looking only at adverse situations (Leveson, 2011; Hollnagel, 2012, 2014a, 2014b).

This paper extends previous research that looked at traces of new safety thinking in investigation reports as a means to detect gaps between knowledge and practice in the field of investigations as well as to examine differences between regions (Karanikas et al., 2015; Karanikas, 2015). Based on the analysis framework presented by Karanikas (2015), in this study we employed a broader set of reports to examine the degree to which the nine aspects of new safety thinking and the three categories of safety models stated in the research mentioned above have been visible in safety investigations published between 1999 and 2016.

2. METHODOLOGY

The framework presented by Karanikas (2015) was converted into an analysis tool (Table 1) with the scope to detect new safety thinking practices (NSTPs) in investigation reports of aviation events and the frequency to which the three safety model types of Table 2 were represented in these reports. It is clarified that the analysis aimed to identify whether each of the aspects was visible at least once in each report. Therefore, we did not adopt a “logistical” approach, meaning that we did not use the tool to count how many times each NTSP of Table 1 could be (not) in each report. The goal was to examine whether there had been efforts during investigations to apply the so-called new approaches to safety and human error.

The analysis tool was designed in the Excel software (Microsoft Corporation, 2013), and its main body includes nine questions (Table 1) each of them corresponding to an NSTP as well as a section with the brief descriptions of the safety model types of Table 2. The analyst should read each report and decide whether each of the questions could be answered as “YES” or “NO” at least once and, in case of a positive answer, to provide respective justifications (i.e. the parts of the report that each NSTP was found). In case that the user of the tool believed that a question was not relevant or traceable to the context of the report, a “NON-APPLICABLE” answer was also available. For example, if human errors had not been mentioned in the investigation report, the fields referring to human error, judgmental attitude and other relevant aspects were scored as non-applicable. Moreover, the analyst was asked to determine the safety model type which was closer to the way the investigation was performed and provide a relevant short justification. Tables 1 and 2 mention the shortcodes used in this paper for the NTSPs and safety model families.

Table 1 Questions used to detect new safety thinking practices (NSTPs)

NSTP	Code	Question used to analyse safety investigation reports
Human error seen as a symptom	HES	Did the investigators search for factors which contributed to the human errors identified?
Hindsight bias minimisation	HBM	Did the investigators follow a forward chronological timeline to explain the choices of the end-users out of the options they had and/or why it made sense to them at that time?
Shared responsibility	SHR	Did the investigators mention various organisational/systemic factors which contributed to the event?
Non-proximal approach	NPA	Did the investigators search the organisational/systemic factors that contributed to the event to the same extent they did for the proximal causes (e.g. human errors of end-users and technical failures)?
Decomposition of folk models	DFM	Did the investigators avoid naming abstracts statements/labels* as causes and try to explain them further? *Abstract statements refer to ideas or concepts which do not have physical referents (e.g., poor communication, lack of awareness, high workload).
Non-counterfactual approach	NCA	Did the investigators try to explain why end-users deviated from standards and procedures or did they examine the applicability of these standards and procedures to the context of the event?
Non-judgmental	NJA	Did the investigators try to explain why end-users deviated from norms and

NSTP	Code	Question used to analyse safety investigation reports
approach		expectations or did they examine the validity of these norms and expectations?
Safety-II	SII	In addition to the failures, did the investigators mention individual, team or organisational/system successes during or before the event or events under similar conditions?
Feedback loops examination	FLE	Did the investigators take the effectiveness of feedback mechanisms* into account? Feedback mechanism: It is a process or component of a system that provides information to another process or component.

Table 2 Brief description of the three safety model families (adapted from Kaspers et al., 2016)

Type	Code	Brief description
Sequential	SEQ	Direct cause-effect relationships: a clearly defined timeline of failures, errors and violations that lead to an event.
Epidemiological	EPD	Direct and indirect cause-effect relationships: clearly defined timeline of active failures along with long-lasting effects of latent problems that contribute to active failures.
Systemic	SYS	Dynamic, emerging and complex system behaviours: examining interactions, interdependencies and relationships between parts to understand a system as a whole, including effects of the behaviour of individual elements.

To finalise the fields of the analysis tool referring to the NSTPs and the safety models, we performed six pilot sessions where we assessed the inter-rater agreement (Bell et al., 2006; Gwet, 2008). The authors, seven students and four external safety experts from the aviation industry participated in different sessions depending on their availability. At the beginning of each session, the running version of the tool was presented and explained. Afterwards, the participants were asked to apply the tool to randomly chosen investigation reports of aviation events. Then, we ran group-focus sessions and discussed any problems regarding the wording, clarity and validity of the questions. Each version of the tool was improved based on the comments of the previous session before executing the next pilot.

In total, 25 different reports were analysed across the six pilot tests. The inter-rater reliability was assessed with the Intra-Class Correlation coefficient test of the SPSS Software version 22 (IBM, 2013) under the settings: two-way mixed, absolute agreement, test value= 0, confidence level 95%. The values of the tests ranged from 0.51 in the early versions of the tool to 0.82 for its current version which was deemed as sufficiently reliable (e.g., Kanyongo et al., 2007). Nonetheless, regardless the achievement of adequate overall reliability of the tool, the discussions after each pilot session suggested that the answers to each of the questions were highly dependable on the knowledge and background of the analyst, possible biases against or in favour of the concepts addressed by new safety thinking and the differences in the wording of the reports. However, the peer-review sessions helped to calibrate the analysts and maintain consistency in the framework's application.

To examine possible variations of the extent to which each NSTP and safety model type had been applied, the tool included fields for the authority which issued the investigation report, the year it was published, actual involvement of end-users into the development of the event (YES/NO) and whether the event resulted in fatalities (YES/NO). The two former fields were adapted based on the practice of industry reports that present data and differences across regions and over time (e.g., IATA, 2018; ICAO, 2018). The two last variables were added as a means to detect variations that might be attributed to easy-to-fix perspectives and the severity/outcome bias (e.g., Evans, 2007; Dekker, 2014; Karanikas & Nederend, 2018). Therefore, the hypotheses tested with the use of the variables mentioned above are the following:

- *HYP1: Over time, there has been an increase of application of all NSTPs during safety investigations.*

The proof of this hypothesis would indirectly justify the use of the term “new” regarding the implementation of the particular aspects and their effective dissemination.

- *HYP2: There are differences amongst regions regarding the extent to which the NSTPs are applied.*

It is expected that new approaches are not embraced by all regions at the same extent due to the effects of different national cultures (e.g., Li & Harris, 2005; Li et al., 2007) which can influence safety management in general.

- *HYP3: The NSTPs have been applied to the same extent regardless of the involvement of the end-users in the development of the event.*

- *HYP4: The NSTPs have been applied to the same extent regardless of the existence of fatalities as a result of the event.*

The hypotheses HYP3 and HYP4 were based on the premise that investigators must be impartial to the maximum degree possible and must be able to manage their feelings, emotions and biases (e.g., Lekberg, 1997; Dekker, 2002).

The safety investigation reports analysed were randomly selected from the online repositories of the Air Accidents Investigation Branch of the United Kingdom, Australian Transport Safety Bureau, Dutch Safety Board, National Transportation Safety Board of the United States and Transportation Safety Board of Canada. The specific authorities were preferred because they publish their reports in the English language and maintain databases of reports for recent and older safety events. Due to time limitations, the maximum number of reports analysed per authority was limited to 60 items maximum, and in total 277 investigation reports published between 1999 and 2016 were processed. It is noticed that the number of reports found on the websites of the particular authorities ranged from 300 to more than 2000 for the specific period, thus the number of reports analysed were not analogous to the ones found on the online repositories. Due to the unrepresentative sample per authority, we could not derive conclusive results per region; therefore, we decided to seal the correspondence between the authorities’ identity and the results by assigning the codes AIAx (x=1-5) randomly.

Table 3 presents the sample size and distribution of the reports across the variables employed in this study. The time of publication was the principal criteria to select and divide the reports (i.e. 2006 and earlier, and 2007 and later). The particular decision was made considering that the communication of new safety thinking commenced mainly after 2004 (e.g., Leveson, 2004; Dekker, 2007) and that the average time between the event dates and the release of their investigation reports for the sample was calculated to two years. The differences in the number of reports processed per authority are due to the different working pace of the students, the different length of the reports per region and severity level as well as the various time length necessary for each student to get familiar with the analysis framework.

Table 3 Distribution of sample

Variables, number of reports (N), and valid percentages (%)			
Authority	Period	End-user involvement	Fatalities
AIA1, N=60, 21.7%	≤ 2006, N=140, 50.5%	YES, N=169, 61.0%	YES, N=99, 35.7%
AIA2, N=45, 16.2%	≥ 2007, N=137, 49.5%	NO, N=108, 39.0%	NO, N=178, 64.3%
AIA3, N=52, 18.7%			
AIA4, N=60, 21.7%			
AIA5, N=60, 21.7%			

In addition to frequency calculations, Chi-square tests were performed to examine possible significant associations of the frequency of application of NSTPs and safety models with the variables mentioned above (i.e. publishing authority, period, end-user involvement and fatalities). Considering the effects of individual interpretations when analysing the reports, as these were evident during the inter-rater agreement tests, the significance level for the statistical tests was set to $\alpha=0.01$ to compensate for subjectivity. We performed all analyses of quantitative data recorded from the reports and surveys in the SPSS Software version 22 (IBM, 2013).

3. RESULTS

The frequencies of the new safety thinking practices (NSTPs) detected at least once in the investigation reports analysed, where applicable, ranged from 26.9% to 79.4% and are presented in Figure 1. *Human error seen as a symptom* (HES), *Decomposition of folk models* (DFM) and *Feedback loops examination* (FLE) were detected at least in three-quarters of the reports. The NSTPs *Hindsight bias minimisation* (HBM), *Shared responsibility* (SHR), *Non-judgmental approach* (NJA) and *Non-counterfactual approach* (NCA) were traced in 50%-75% of the cases, whereas *Non-proximal approach* (NPA) and *Safety-II* (SII) were the least represented aspects. Regarding the safety model types, the Epidemiological one was found in 52.7% of the reports, the Sequential one was detected in 44.1% of the cases and in the rest 3.2% of the reports a systemic model was followed.

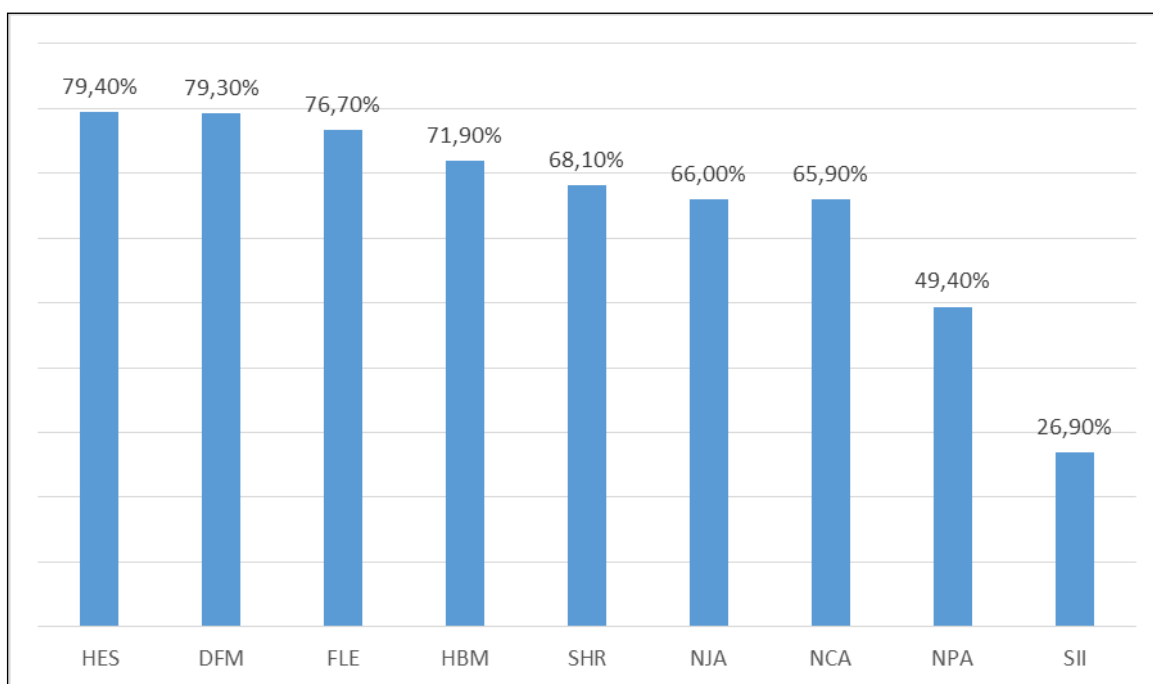


Figure 1: Frequencies of NSTPs applied to investigations

The results of the statistical tests are presented in the Tables 4 and 5. The former table shows the differences amongst authorities and the latter reports the variances for the rest of the variables (i.e. period, end-user involvement and fatalities). It is noticed that we excluded systemic models from the statistical calculations due to the low number of reports in which they were detected. The results indicated that the frequencies to which the NSTPs had been applied were quite different across the regions included in the study, such differences being significant for the HBM, SHR, NPA, SII practices as well as the distribution between Sequential and Epidemiological models. The AIA5 was found with the least application frequency for all four practices mentioned above with the values ranging from 13.3% to 55%. The highest percentage of application for the particular practices was detected in AIA2 reports for HBM (93.3%), NPA (75.0%), and SII (53.3%)

and AIA4 reports for SHR (85%). Regarding the safety model type, AIA1 had applied Epidemiological models with the highest frequency (75.9%), and AIA4 had the highest percentage of Sequential models' application (63.8%).

Table 4 Results of statistical tests for the authorities

NSTP (N)	Authority (% of cases in which the aspect was traced)					p-value*
	AIA1	AIA2	AIA3	AIA4	AIA5	
Human error seen as symptom (N=194)	95.5	77.8	69.4	73.5	78.9	p=0.037
Hindsight bias minimisation (N=260)	78.3	93.3	45.7	81.7	55.0	<u>p=0.000</u>
Shared responsibility (N=263)	75.0	76.9	59.1	85.0	45.0	<u>p=0.000</u>
Non-proximal approach (N=261)	38.3	75.0	53.7	70.0	20.0	<u>p=0.000</u>
Decomposition of folk models (N=251)	76.7	87.5	78.7	70.0	88.3	p=0.117
Non-counterfactual approach (N=208)	65.9	81.8	65.8	63.8	50.0	p=0.54
Non-judgmental approach (N=212)	59.6	86.4	63.2	59.6	61.5	p=0.34
Safety-II (N=275)	28.3	53.3	14.0	30.0	13.3	<u>p=0.000</u>
Feedback loops examination (N=275)	86.7	75.6	72.0	80.0	68.3	p=0.152
Safety model family (N=268)	Distribution of model types across the cases (%)					p-value*
	AIA1	AIA2	AIA3	AIA4	AIA5	
Sequential	24.1	31.1	49.0	63.8	55.0	<u>p=0.000</u>
Epidemiological	75.9	68.9	51.0	36.2	45.0	

* Statistically significant results underlined

An observation of the results regarding the period (Table 5) suggests that all NSTPs were identified more frequently from 2007 and later. However, the differences were statistically significant only for *Safety-II* with an increase of about 15% in the second period. When the end-user was involved, there had been no significant variances. However, it was observed that most of the NSTPs were slightly more frequently applied when there was no direct involvement of the end-user in the event. In the case of the fatalities variable, a significant variation was detected only for the *Feedback loop examination*, where the specific practice was applied at a lower extent when the event had resulted in casualties.

Table 5 Results of statistical tests for the variables of period, end-user involvement and fatalities

NSTP (N)	Variables (% of cases in which the aspect was traced)								
	Time period			End-user involvement			Fatalities		
	≤ 2006	≥ 2007	p value*	Yes	No	p value*	Yes	No	p value*
Human error seen as symptom (N=194)	74.2	84.5	p=0.076	78.4	84.4	p=0.445	77.8	80.5	p=0.640
Hindsight bias minimisation (N=260)	70.2	73.6	p=0.540	68.9	77.4	p=0.141	74.5	70.4	p=0.474
Shared responsibility (N=263)	63.9	72.3	p=0.174	67.5	69.1	p=0.778	66.7	68.9	p=0.706
Non-proximal approach (N=261)	44.4	54.7	p=0.095	49.1	50.0	p=0.891	47.5	50.6	p=0.622
Decomposition of folk models (N=251)	74.4	84.4	p=0.051	78.2	81.4	p=0.551	76.3	81.2	p=0.353
Non-counterfactual approach (N=208)	60.6	71.7	p=0.090	63.8	71.2	p=0.308	62.2	67.9	p=0.403
Non-judgmental approach (N=212)	62.2	70.3	p=0.212	61.5	75.4	p=0.046	58.1	70.3	p=0.074
Safety-II (N=275)	19.4	34.6	<u>p=0.005</u>	22.5	34.0	p=0.037	21.2	30.1	p=0.110
Feedback loops examination (N=275)	72.5	81.0	p=0.093	79.3	72.6	p=0.204	67.7	81.8	<u>p=0.008</u>
Safety model type (N=268)	Distribution of model types across the variable values (%)								
Sequential	45.9	45.1	p=0.894	41.6	52.0	p=0.097	48.5	43.9	p=0.468

	Time period		End-user involvement		Fatalities	
	≤ 2006	≥ 2007	Yes	No	Yes	No
	Epidemiological	<u>54.1</u>	<u>54.9</u>	<u>58.4</u>	<u>48.0</u>	<u>51.5</u>

* Statistically significant results underlined

4. DISCUSSION

The overall results indicate that all new safety thinking practices were more or less visible across the whole sample, but with remarkable differences in their frequencies. First, it seems that investigation teams had sufficiently embraced the concept that the detection of human error cannot constitute the end point of investigations. Although the researchers did not aim to examine whether this approach was followed in all the cases where unsuccessful human interventions were stated in each report, the detection of this aspect in almost 80% of the sample can be attributed to the fact that the “human error seen as a symptom” perspective has been advocated long before recent literature was published. For example, the popular Swiss Cheese Model (Reason, 1990) and its extension through the Human Factors Analysis and Classification System (HFACS) by and Wiegmann and Shappell (2003) had already communicated the contribution of latent factors that influence decision making and deeds of end-users. Moreover, we cannot neglect the fact that investigators are not isolated from their experiences in the real-world arena. From that perspective, their efforts to look behind human error might stem naturally from a reflection on unpleasant situations they confronted and, possibly, their will to protect others from the unfair treatment of their actions and decisions as the final “root-causes” of unfavourable events. Nonetheless, almost one fifth of the reports had stopped at the attribution of human error as the final cause of events; this result suggests that there might be plenty of occasions where investigations focus on the performance of the end-users alone and, apart from laying the ground for a blaming culture, they deprive systems of a more profound learning potential.

Regarding the “decomposition of folk models” and “feedback loop examination” that were recorded in more than 75% of the cases studied, the authors believe that the prevalence of engineers in the safety investigations field and, expectedly, their familiarity with systems engineering have led them to (1) avoid the labelling of constructs as event causes because they search for measurable and observable causal factors, and, (2) examine whether systems provided to the end-users with adequate information about the state and outcomes of processes (e.g., instrumentation). The 25% of the reports that we did not find relevant attempts show that besides any general investigation constraints (e.g., budget, time, resources), constructs might be misused as the new scapegoats or function as bases for generalisation of results instead of an opportunity to dive deeper in their underlying mechanisms. The lack of reference to feedback loops’ existence or effectiveness might be the result of investigators biases such as the availability heuristic, through which judgement is affected by past frequency and gravity to personal experience (Tversky & Kahneman, 1974; Greene & Ellis, 2007) (e.g., an investigator might have flown the same aircraft type and had not experienced any problem with the information provided by the system), or anchoring, which inhibits judgment due to anchors such as salient numerical values or other pieces of information (Tversky & Kahneman, 1974; de Wilde et al., 2018) (e.g., an investigator might evaluate cognitive effects only by examining recent events).

The detection of “hindsight bias minimisation” in about 72% of the reports is seen as a highly positive sign. Hindsight bias is quite difficult to overcome due to the effects of the confirmation bias and the inevitably backwards process of the investigation. Regarding the former, although scientific principles demand from researchers to test and not just confirm their findings (Kassin et al., 2013; Fforde, 2017), this might not always be the case in real-world practice. The latter factor is linked to

the fact that the starting point of all investigations is the collection of facts and evidence from the incident/accident scene. Following this phase, investigators reconstruct the development of the event in a backwards direction; it is practically impossible for investigation teams to know beforehand how the events unfolded. However, it seems that investigators might not comprehend that a backwards research will merely uncover eventualities (i.e. what happened) and lead to possible reasons (i.e. why happened), whereas a more complete explanation of any occurrence (i.e. how happened) is only feasible if a research along a forward timeline accompanies the backwards examination.

The appearance of “shared responsibility” in about 68% of the reports can be ascribed to the reasons stated above regarding the “human error seen as symptom” aspect which, though, was detected at a quite higher frequency. This difference, on the one hand, can be attributed to the fact the latent problems influencing human performance do not only refer to human agents but include technology-related factors, which literally cannot carry any responsibility. On the other hand, the particular aspect might not have been very frequently applied due to difficulties to gain access to persons serving in different organisations or different hierarchical levels within the same organisation, possible self-imposed inhibitions of investigators (e.g., questioning the actions and decisions of ex-colleagues and senior staff), diverse interpretations of work ethics (e.g., the more people involved in the list of findings, the worse for them and the organisation) or implicit and explicit constraints imposed by higher organisational and system levels.

The attempts to follow the “non-judgmental” and “non-counterfactual” approaches were recorded in about 66% of the sample. Although these figures indicate that investigators put efforts to examine the applicability and validity of standards and expectations within the context of each occurrence, the non-detection of these aspects in one third of the reports suggests that compliance with procedures was seen in many cases as a sufficient and necessary condition to deal with the dynamic and complex operating environment. This situation possibly signals the effects of biases such the ones described above for the “feedback loops examination” aspect as well as effects of regional safety management culture regarding the examined sample of reports. The appearance of the “non-proximal approach” in about 50% of the reports with a 20% difference from the “shared responsibility” aspect, in addition to psychological, emotional and managerial influences discussed above, might be attributed to the relative easiness to trace information and data regarding the lower level of operations. Although in an ideal situation each system function that contributed to the event should be thoroughly examined, it is not atypical that the decisions made and the actions performed long before the event cannot be evidently traced or explained (e.g., not logged/documented, documented but without the underlying reasoning, unavailability of persons involved, decayed memory of witnesses approached).

Rather expectedly the “safety-II” aspect was the one with the lowest representation in the sample analysed. This extremely low frequency can be attributed to the prevalent practice of addressing failures as part of the learning process (Madsen & Desai, 2010) and the focus of safety and accident models on system failures (Hollnagel et al., 2015). The model families concerned, the rare application of systemic models could be justified by their relatively recent publication and dissemination along with the time needed to gain grounds in industry practice. However, considering that epidemiological models had been already introduced since the 90’s, it would be expected that their application would exceed the application of sequential models significantly. Apart from possible resource constraints, this phenomenon might indicate a low adaptation of safety training and education material, the resistance of established investigation practices to newer approaches or lack of proper background, even motivation, of the various stakeholders who receive the investigation reports. In business practice, it is not rare that managerial levels seek for brief and inclusive summaries of studies as an initial crisis communication response to stakeholders (Coombs & Holladay, 2010). This practice might urge investigators to adopt a more

simplistic approach to explain and present occurrences, thus preferring sequential models over epidemiological models, and the latter over systemic ones.

Our first hypothesis was confirmed partially. Although after 2007 there was a higher frequency of all NSTPs in safety investigation reports, only the “safety-II” aspect was found statistically significant. The overall increase of frequencies is aligned with our earlier argument that NSTPs were communicated more broadly after 2004, while the significant increase of the “safety-II” approach can be attributed to its novelty. As discussed above, the appreciation of successes in addition to the study of failures is not a well-established concept of daily practice. The non-significant increase of the rest of the NSTPs might be explained by the fact that the particular aspects had been already more or less, informally and formally, part of investigation practices along with the influence of similar perspectives on human error published decades ago (Heinrich, 1931; Reason, 1990). Nonetheless, other factors that affect the investigation process (e.g., inadequate resources, external pressures, personal predispositions, content and quality of respective educational programmes) might have also played a role in limiting the more frequent application of the NSTPs over time.

The hypothesis regarding the significant differences in the degree of application of the NSTPs across regions was confirmed for four aspects and the family models; this finding partially confirms the effects of different national cultures (e.g., Li & Harris, 2005; Li et al., 2007). Although due to the lack of relevant empirical research the authors cannot state plausible explanations for the observation of differences only in half of the aspects studied, it is crucial to notice that the existence of variations indicates a lack of a harmonised approach to investigations. Indeed, the researchers do not support conformity to any “thinking standard”, but the diversity of approaches to safety can hide a different treatment of individuals depending on where the accident occurred [i.e. in the aviation industry, in principle the State of occurrence initiates and leads the investigation (ICAO, 2001)].

Additionally, the slight increase of application of most of the NSTPs when there was no direct involvement of the end-user into the event suggests that cases of the latter might trigger the emergence of investigation biases (e.g., Lekberg, 1997; Dekker, 2002) and lead to simplified and less profound findings. Moreover, all but the “feedback loops examination” practice had been implemented to the same extent irrespective of the existence of fatalities as a result of the events. The authors believe that the overall picture shows a well-mobilised professional approach of investigation teams where the emotional influences due to fatal injuries had been sufficiently controlled. The variation recorded in the examination of feedback loops might be explained by the lack of respective information and data from the end-users that were victims.

5. CONCLUSIONS

This research presented the results of the application of an analysis framework, which reflects nine new safety thinking practices and three safety model categories, to a large sample of safety investigation reports. The aim of the study was not to count how many times each new safety-related approach was applied to each case analysed but to gain initial insights regarding attempts to implement new safety thinking over time and across regions. The overall results showing that all NSTPs were visible during investigations at various degrees, on the one hand, suggests that these practices are relevant to investigators and not completely unknown. On the other hand, the variety of frequencies of implementation per NSTP were attributed by the authors to the individual, organisational and systemic factors and constraints that might influence investigation professionals. However, further research is necessary to examine (1) whether the implementation of the NSTPs employed in this research stems from knowledge gained through training, education and self-development efforts (e.g., reading respective literature) or practices evolved and exchanged amongst professionals, and (2) factors that can be supportive or opposing

to new safety thinking. Also, the application of the analysis framework to larger and more representative samples from aviation and other industry sectors is expected to offer further insights within and across various domains. We strongly recommend to perform a consistency check amongst analysts before any study (e.g., peer-reviews), as practised in this research, to minimise biases that would lead to the (non)detection of NSTP aspects and unreliable results.

As a last consideration, the statistical results showed a non-significant variation over time of 8 out of the 9 NSTPs for the period 1999-2016. This finding might be an indication that the so-called new safety thinking practices might have been part of investigation practice long before recent literature focused immensely on a human-centric and systems approach. Nonetheless, we should not underestimate the impact of contemporary researchers and authors and their efforts to foster further the respective concepts, even if we accept that investigators had been already familiar with these at different extents. It can be confidently argued that NSTPs offer the opportunity for fairer and more in-depth analyses. However, since investigations are subject to boundaries, even these are not always explicitly recognised and reported (Plioutsias et al., 2018), we suggest the communication of any modern practices in the direction of acquiring a deeper understanding of causality and learning opportunities for the whole system and not judging investigation practices without considering the overall investigation limitations and constraints.

REFERENCES

- Bell, V., Halligan, P. W., & Ellis, H. D. (2006). Diagnosing Delusions: A Review of Inter-Rater Reliability. *Schizophrenia Research*, 86, 76-79. doi:10.1016/j.schres.2006.06.025
- Coombs, T. W., & Holladay, S. J. (2010). *The Handbook of Crisis Communication*. USA: WILEY-BLACKWELL Publishing Ltd.
- de Wilde, T. R., Ten Velden, F. S., & De Dreu, C. K. (2018). The Anchoring-Bias in Groups. *Journal of Experimental Social Psychology*, 116-126. doi:10.1016/j.jesp.2018.02.001
- Dekker, S. W. (2002). Reconstructing Human Contributions to Accidents: The New View on Error and Performance. *Journal of Safety Research*, 33, 371-385.
- Dekker, S. W. (2007). Six Stages To The New View of Human Error. *Safety Monitor*, 11(1), 1-5. Retrieved July 23, 2018, from https://www.researchgate.net/publication/255591192_Six_stages_to_the_new_view_of_human_error
- Dekker, S. W. (2014). The Field Guide to Understanding Human Error. In S. Dekker, *The Field Guide to Understanding Human Error*. Surrey, United Kingdom: Ashgate Publishing Limited.
- Evans, J. K. (2007). *An Application of CICCT Categories to Aviation Accidents in 1988-2004*. NASA/CR-2007-214888. Hanover, MD: US National Aeronautics and Space Administration.
- Fforde, A. (2017). Confirmation Bias: Methodological Causes and a Palliative Response. *Quality and Quantity*, 51(3), 2319-2335. doi:10.1007/s11135-016-0389-z
- Greene, E., & Ellis, L. (2007). Decision Making in Criminal Justice. In D. Carson, B. Milne, F. Pakes, K. Shalev, & A. Shawyer, *Applying Psychology to Criminal Justice* (pp. 183-200). Chichester: Wiley. doi:10.1002/9780470713068.ch11
- Gwet, K. L. (2008). Computing Inter-Rater Reliability and Its Variance in the Presence of High Agreement. *British Journal of Mathematical and Statistical Psychology*, 61, 29-48. doi:10.1348/000711006X126600
- Heinrich, H. W. (1931). *Industrial Accident Prevention: A Scientific Approach* (1st ed.). New York: McGraw-Hill.
- Hollnagel, E. (2012). *FRAM, The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*. Ashgate Publishing, Ltd.
- Hollnagel, E. (2013). A Tale of Two Safeties. *Nuclear Safety and Simulation*, 4(1), 1-9. Retrieved July 23, 2018, from <http://erikhollnagel.com/A%20Tale%20of%20Two%20Safeties.pdf>
- Hollnagel, E. (2014a). Is Safety a Subject for Science? *Safety Science*, 67, 21-24. doi:10.1016/j.ssci.2013.07.025

- Hollnagel, E. (2014b). *Safety-I and Safety-II: The Past and Future of Safety Management*. Farnham, UK: Ashgate Publishing Limited.
- Hollnagel, E., Wears, R. L., & Braithwaite, J. (2015). *From Safety-I to Safety-II: A White Paper*. University of Southern Denmark, University of Florida, USA, Macquarie University, Australia.
- Hollnagel, E., Woods, D. D., & Leveson, N. C. (2006). *Resilience Engineering: Concepts and Precepts*. Aldershot, UK: Ashgate.
- IBM. (2013). IBM SPSS Statistics for Windows version 22. Armonk, NY: IBM Corp.
- International Air Transport Association-IATA. (2018). *Safety Report 2017* (54th ed.). Montreal: International Air Transport Association. Retrieved September 22, 2018, from <https://aviation-safety.net/airlinesafety/industry/reports/IATA-safety-report-2017.pdf>
- International Civil Aviation Organization-ICAO. (2001). *Aircraft Accident and Incident Investigation - Annex 13*. ICAO. Retrieved September 19, 2018, from https://www.emsa.europa.eu/retro/Docs/marine_casualties/annex_13.pdf
- International Civil Aviation Organization-ICAO. (2018). *Safety Report*. Montreal: International Civil Aviation Organization. Retrieved September 22, 2018, from https://www.icao.int/safety/Documents/ICAO_SR_2018_30082018.pdf
- Kanyongo, G. Y., Brook, G. P., Kyei-Blankson, L., & Gocmen, G. (2007). Reliability and Statistical Power: How Measurement Fallibility Affects Power and Required Sample Sizes for Several Parametric and Nonparametric Statistics. *Journal of Modern Applied Statistical Methods*, 6(1), 81-90. doi:10.22237/jmasm/1177992480
- Karanikas, N. (2015). Human Error Views: A Framework for Benchmarking Organizations and Measuring the Distance between Academia and Industry. *Proceedings of the 49th ESReDA Seminar, 29-30 October 2015*. Brussels, Belgium.
- Karanikas, N., & Nederend, J. (2018). The Controllability Classification of Safety Events and Its Application to Aviation Investigation Reports. *Safety Science*, 108, 89-103. doi:10.1016/j.ssci.2018.04.025
- Karanikas, N., Soltani, P., de Boer, R. J., & Roelen, A. (2015). Evaluating Advancements in Accident Investigations Using a Novel Framework. *Proceedings of the 5th Air Transport and Operations Symposium (ATOS), 20-22 July 2015*. Delft University of Technology, Netherlands.
- Kaspers, S., Karanikas, N., Roelen, A. C., Piric, S., & de Boer, R. J. (2016). *Review of Existing Aviation Safety Metrics, RAAK PRO Project: Measuring Safety in Aviation, Project number: S10931*. The Netherlands: Aviation Academy, Amsterdam University of Applied Sciences.
- Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions. *Journal of Applied Research in Memory and Cognition*, 42-52. doi:10.1016/j.jarmac.2013.01.001
- Lekberg, A. K. (1997). Different Approaches to Incident Investigation-How the Analyst Makes. *Hazard Prevention*, 33(4), 10-13.
- Leveson, N. G. (2004). A New Accident Model for Engineering Safer Systems. *Safety Science*, 42(4), 237-270. doi:10.1016/S0925-7535(03)00047-X
- Leveson, N. G. (2011). Applying Systems Thinking to Analyse and Learn from Events. *Safety Science*, 49, 55-64. doi:10.1016/j.ssci.2009.12.021
- Li, W. C., & Harris, D. (2005). HFACS Analysis of ROC Air Force Aviation Accidents: Reliability Analysis and Cross-Cultural Comparison. *International Journal of Aviation Studies*, 5(1), 65-81.
- Li, W. C., Harris, D., & Chen, A. (2007). Eastern Minds in Western Cockpits: Meta-Analysis of Human Factors in Mishaps from Three Nations. *Aviation, Space, Environmental Medicine*, 78(4), 420-425.
- Madsen, P. M., & Desai, V. (2010). Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry. *Academy of Management Journal*, 53(3), 451-476. doi:10.5465/amj.2010.51467631
- Martinetti, A., Chatzimichailidou, M. M., Maida, L., & van Dongen, L. (2018). Safety I-II, Resilience and Antifragility Engineering: A Debate Explained through an Accident Occurring on a Mobile

- Elevating Work Platform. *International Journal of Occupational Safety and Ergonomics*. doi:10.1080/10803548.2018.1444724
- Microsoft Corporation. (2013). Microsoft Excel 2013. Santa Roza, California: Microsoft Office Professional Plus 2013.
- Plioutsias, A. & Karanikas, N. & Tselios, D. (2018). Decreasing the Distance Between International Standards from Different Domains: The Case of Project Management and Aviation Safety Investigations, *Proceedings of the International Cross-industry Safety Conference, 2-3 November 2017*, Aviation Academy, Amsterdam University of Applied Sciences, AUP Advances, 1(1), pp. 7-39, doi: 10.5117/ADV2018.1.002.PLIO
- Reason, J. (1990). *Human Error*. New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.
- Wiegmann, D. A., & Shappell, S. A. (2003). *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Burlington, VT: Ashgate Publishing, Ltd.