Amsterdam University of Applied Sciences

The controllability classification of safety events and its application to aviation investigation reports

Karanikas, Nektarios; Nederend, Jeffrey

Link to publication

The Controllability Classification of Safety Events and its Application to Aviation Investigation Reports

Abstract

This paper proposes an amendment of the classification of safety events based on their controllability and contemplates the potential of an event to escalate into higher severity classes. It considers (1) whether the end-user had the opportunity to intervene into the course of an event, (2) the level of end-user familiarity with the situation, and (3) the positive or negative effects of end-user intervention against expected outcomes. To examine its potential, we applied the refined classification to 296 aviation safety investigation reports. The results suggested that pilots controlled only three-quarters of the occurrences, more than three-thirds of the controlled cases regarded fairly unfamiliar situations, and the flight crews succeeded to mitigate the possible negative consequences of events in about 71% of the cases. Further statistical tests showed that the controllability-related characteristics of events had not significantly changed over time, and they varied across regions, aircraft, operational and event characteristics, as well as when fatigue had contributed to the occurrences. Overall, the findings demonstrated the value of using the controllability classification before considering the actual outcomes of events as means to support the identification of system resilience and successes. The classification can also be embedded in voluntary reporting systems to allow end-users to express the degree of each of the controllability characteristics so that management can monitor them over time and perform internal and external benchmarking. The mandatory reports concerned, the classification could function as a decision-making parameter for prioritising incident investigations.

Keywords: controllability; event classification; event severity; event potential; system resilience

1. Introduction

Despite the continuous increase in aviation safety levels over the past half-century (Boeing, 2017; Airbus, 2017), additional efforts are put to improve safety further by monitoring safety performance through respective indicators (Bellamy & Sol, 2012; Kjellén, 2009; Verstraeten, Roelen & Speijker, 2014). Regulations, standards and industry practice dictate the classification of safety events based on their actual severity. The use of event rates (e.g., number of accidents per unit of activity/exposure) prevails, thus suggesting a focus on recorded consequences to demonstrate safety performance (e.g., Airbus, 2017; Boeing, 2017; HSE, 2016; EASA, 2016a, ICAO, 2013a).

Amongst the various safety management activities that aim to improve safety by preventing reoccurrence of adverse events, States and organisations are required to conduct investigations of accidents and serious incidents (ICAO, 2010; EU, 2010). Despite the argument that near-misses might remain unreported and unrecorded unless their effects cannot be hidden (Bhagwati, 2006), the enforcement of mandatory reporting systems allows the recording of various characteristics of safety events. The collection of such data enables the industry and authorities to perform statistics, analyse associated factors and monitor trends. Although the conduction of investigations for incidents is non-obligatory, it is recognised that those also comprise opportunities for obtaining information that can lead to an increase of safety levels (ICAO, 1993; Wise, Hopkin & Garland, 2009). However, under the reality of limited resources, those are mostly devoted to the investigation of serious incidents and accidents (ICAO, 2015; Wise, Hopkin & Garland, 2009; Greenwell, 2003). Furthermore, risk levels are detected and prioritised using such an outcome-oriented approach (e.g., EASA, 2016b), which also

prevails the threshold between voluntary and mandatory safety reporting. In general, current views on aviation safety and associated improvement initiatives concentrate principally on the severity of reported or anticipated events, which informs the decisions about focus areas and allocation of respective resources.

By considering the role of the human element in the development of events, Karanikas (2015) introduced a new classification scheme that incorporates the potential of an occurrence to escalate instead of counting only for its actual outcome(s). The author above contemplated that a sole emphasis on outcomes does not address the extent to which the system users control events. Karanikas (2015) showed that it is important to examine whether the outcome of an occurrence was associated with an attempt or opportunity to control an unfolding situation and consider the effectiveness of human interventions to alleviate the possible consequences of the safety event. The application of the suggested classification on a large aviation organisation concluded that (1) the specific classification scheme might function as an additional or alternative measurement of safety performance before focusing on the severity of eventualities, and (2) various factors were associated with the controllability of events, such as aircraft type and generation, and operating unit.

The primary goal of the study presented in this paper was to apply the classification of Karanikas (2015) to a sample of safety investigation reports published by various aviation authorities as a means to examine the value of the classification at a wider context and trigger the interest for its application by various industry sectors. Furthermore, this research aimed at evaluating the classification's potential to serve as a safety performance metric, as opposed to dominant severity-based metrics, and supplement the safety perspective of organisations and States by exploring associations of this classification with event characteristics and factors. Finally, based on the findings of this study, the authors reflected on the potential of the controllability classification to support the prioritisation of incident investigations and its connection with modern safety thinking and initiatives, such as system resilience and Safety-II.

The work presented in this paper is organised as follows. In section 2, we present different types of safety event classifications and discuss various factors and characteristics of events discussed in studies and industry reports. Section 3 describes the reasons that led to the amendment of the controllability classification and is followed by the methods and materials used in its application and the analysis of data. The results of the research are presented in section 4 of the paper and discussed in section 5 against literature along with relevant limitations of the study. Finally, in the conclusions (Section 6), we present the value of the current study concerning various options for its application, the overall picture from the data analysis as well as recommendations for future research.

2. Literature review

2.1 Classification of safety events

Safety events, irrespective of industry sector, are commonly classified according to the magnitude of their actual impacts on the environment, infrastructure and equipment, and infliction of injuries or casualties. In general, regulatory bodies utilise their classifications of safety occurrences to depict safety performance through accident frequencies and rates (e.g., EMSA, 2016; ICAO, 2017b; EASA, 2016) and decide whether they will launch a safety investigation. Each classification indicates the threshold above which safety investigations are obligatory or to be conducted in the interest of the respective national investigation body. Investigation agencies shall perform safety investigations for accidents, serious accidents, and very serious casualties. Despite being more abundant, less severe

occurrences are investigated in cases of retrievable data and their potential to lead to safety improvements (e.g., EU, 2009; EU, 2010; IMO, 1997; EUAR, 2016).

The classification categories of safety events vary across industry sectors. For example, events in the railway domain are categorised as serious accidents, accidents and incidents (EU, 2016). In the maritime industry, an event can be named as a marine casualty, serious casualty, very serious casualty or marine incident (IMO, 1997). The classification of occurrences in the aviation industry emphasises the potential of the event to develop into an accident with severe consequences. A serious incident is an occurrence that had a high potential to escalate into an accident (ICAO, 2010; EU, 2010), and its classification depends on the analyst's interpretation (Greenwell, 2003). The indicative examples given above apart from the different category names regard also diverse thresholds between the categories. The aviation sector concerned, Kaspers, Karanikas, Roelen, Piric, van Aalst & de Boer (2016) discussed that the ambiguity in standards regarding the threshold between serious incidents and incidents might lead to diverse interpretations and, therefore, render rates or frequencies of events other than accidents as an unreliable safety performance metric. Overall, apart from the definitions of accidents that in all industry domains include the case of fatal injuries and almost catastrophic implications on other assets, the existence of various severity classifications across the industry does not allow a reliable comparison between them regarding rates of events other than accidents. Consequently, accident rates remain the principal indicator used for benchmarking amongst organisations and industry sectors.

The impact-based classification of events is also used in risk management, where analysts rank safety occurrences according to the level of their expected consequences and probability. The former parameter is estimated qualitatively according to the actual severity of similar past events and is complemented with expert judgment. Probabilities can be derived with either quantitative methods when adequate and reliable data are available or a qualitative approach based on the frequency of similar events in the past; in such evaluations, the engagement of experts remains as an option. The two parameters mentioned above are crossed in a respective matrix, and the risk level of an event is determined with the scope to inform decision-making for allocating resources to control risks of higher rank (e.g., ICAO, 2013a; IMO, 2015; EC, 1996; Stamatelatos & Dezfuli, 2011). However, the lack of standardisation of matrices across and within industry sectors, the inherent ambiguities in the categories of severity and likelihood, and the cognitive biases affecting expert judgment threaten the validity and reliability of such an approach (Hubbard & Evans, 2010; Duijm, 2015; Karanikas & Kaspers, 2016).

In the Air Traffic Management (ATM) domain, for instance, Eurocontrol (2009a) has mandated its member States to comply with the "European Safety Regulatory Requirements" (ESARRs). ESARR details the assessment and reporting of events based on a defined list of ATM-related occurrences, divided into accidents and incidents, which, as a minimum, each State report and evaluate. Also, ESARR define the safety data to be communicated with Eurocontrol to identify key risk areas to improve overall operational safety in the ATM system (e.g., rates of occurrences or flight deviations). Having realised that standard metrics of safety rates and traffic volume alone do not sufficiently represent the overall system-wide performance, Eurocontrol (2009b) introduced the Aerospace Performance Factor (APF). The particular metric aggregates various factors related to operational safety risks retrieved from reported incidents and uses a time-variant value that demonstrates the overall risk and performance trend over time as a means to foster safety proactively. To infrom decision-making, safety performance is measured through the APF based on a substantive set of safety metrics, risk assessments from experts and its normalisation against overall traffic volumes. The methodology described above focusses mainly on actual or possible deviations from expected

performance that can lead to more severe events. Thus it constitutes an outcome-based assessment of potential harm which informs organisational decisions (Di Gravio, Mancini, Patriarca, & Costantino, 2015).

In his work, Karanikas (2015) highlighted the outcome bias that prevails the industry in safety performance metrics which do not consider an event's potential to escalate or the efforts of involved personnel to alleviate the anticipated event's consequences. A new classification was therefore suggested based on the controllability of safety occurrences with the intent to differentiate between events with and without user's intervention and indicate the effectiveness of actions of involved personnel to mitigate the ultimate outcomes (Table 1). It is noted that the author used the term 'accident' to refer to safety occurrences of all severity levels used by the particular organisation. According to Karanikas (2015), the potential of an event to escalate into an occurrence of higher severity class is linked to the opportunity and attempt of humans to intervene and control the event. The end-state of a system after an event could be the result of a controlled intervention or an uncontrolled situation, such an intervention being characterised by the desired or undesired influence on the outcome(s) of the event. In addition to the portrayed contrast between controlled and uncontrolled events, a neutral event category enabled the classification of inevitable or expected outcomes of the user's actions in reference with prescribed reactions and application of normal procedures. The particular classification was applied to 808 events of a single organisation that occurred within an eleven-year period. Several factors (e.g., time, aircraft type/generation and operating base characteristics) were considered to study their potential association(s) with the controllability variables. The results of the specific study suggested that the classification could function as a more realistic way for any industry to measure organisational safety performance before considering actual severities (Karanikas, 2015).

Accident control classification	User reaction classification
Controlled: The user attempted to control the accident march	Positive: User's actions did not worsen the outcome; the accident outcome was managed successfully; no errors or violations were noticed during the control attempt
	Negative: User's actions following the safety event initiation resulted in adverse outcomes due to human errors or violations
Uncontrolled: Safety event's consequences were developed without control; there had been no intervention until the time the outcomes were noticed.	None
Neutral: Inevitable application of normal procedures; standard reactions to identified problem	As expected by prescribed procedures

Table 1: Accident control classification (Karanikas, 2015)

2.2. Factors and characteristics of aviation safety events

In addition to severity, various characteristics of safety events are used to identify common issues and steer local, regional or global efforts towards the control of the riskiest/weakest areas. For instance, in the aviation domain, ICAO (2017a) developed the Accident/Incident Data REPorting (ADREP) taxonomy according to which safety events can fall in multiple discrete categories (event type, event phase, aircraft category, etc.). This taxonomy enables the aggregation of data, and the exploration of associations within the categories included or other parameters. An example of the implementation and utilisation of such a categorisation is the "high risk [occurrence] categories" of *Loss Off Control in*

Flight, *Controlled Flight Into Terrain* and *Runway Safety* event types, which constitute the most frequent accident types and have been viewed as the industry's safety priority (ICAO, 2017b). The literature cited in this section presents the main safety event factors and characteristics included in older and more recent studies and industry reports.

Albeit accident and fatality rates in the aviation industry have declined over the last decades (Allianz, 2014; Airbus, 2017; Boeing, 2017), human error and performance remain amongst the most discussed factors in accident causation. As Baker, Qiang, Rebok & Li (2008) and Li, Baker, Grabowski & Rebok (2001) presented, the rates of mishaps due to pilot error declined between 1983 and 2002 at a range of 20% to 40% for commercial air traffic, whereas such rates remained constant for commuter, air taxi and general aviation flight operations. However, the type of errors seemed unchanged for both commercial and general aviation in the US region (Shappell, Detwiler, Holcomb, Hackworth, Boquet & Wiegmann, 2006; Shappell & Wiegmann, 2003a). In older studies, O'Hare, Wiggins, Batt & Morrison (1994) and Wiegmann & Shappell (1997) showed that the magnitude of outcomes of aviation safety events, as depicted by the level of injury and/or damage, corresponded to different error types. Strategic decision errors and failures to set a correct goal to resolve hazardous conditions resulted more often in severe outcomes, while less harmful outcomes were more often connected with failures in executing proper procedures. When comparing fatal against non-fatal accidents, Shappell et al. (2006) and Wiegmann & Shappell (2003) showed that violations of rules and regulations were three times more likely to be associated with fatal occurrences than with non-fatal events. Nevertheless, a focus on human failures has prevailed respective studies, and the cases when end users succeeded under adverse conditions have not been sufficiently analysed to learn from positive results of human performance (Hollnagel, 2014). Also, the analysis of human ability to recover from hazardous conditions and successfully intervene in accidents' trajectory is rarely addressed (Sarter & Alexander, 2000; Reason, 2008).

Regarding the occurrence of safety events in different geographical locations, accident data analysis has led to findings of human performance variance over time for specific countries (e.g., Baker, Qiang, Rebok & Li, 2008; Gaur, 2005; Li, Harris & Yu, 2008), between regions (e.g. Detwiler, Hackworth, Holcomb, Boquet, Pfleiderer, Wiegmann & Shappell, 2006) or across different national cultures (e.g. Li, Harris & Chen, 2007; Li & Harris, 2005). Overall, significant differences in accident rates across regions of occurrence have been recorded. For example, the International Air Transport Association (IATA, 2017) reported the accident rates of commercial civil aviation across various global regions. The Middle East and North Africa, Commonwealth of Independent States, Latin America and the Caribbean obtained in 2016 the highest accident rates per million flight sectors, and, in descending order, the accident rate of Africa was followed by Asia Pacific, Europe, North America and North Asia.

Concerning aircraft generations, Airbus (2017) argued that significant improvements in accident rates have been achieved due to advancements in cockpit technology installed in newer aircraft (e.g., glass cockpit and flight envelope protection). However, increased automation in the cockpit shifted the direct handling role of aircrews into a more detached, monitoring function, and, as a result, pilots intervene mainly when they notice an unanticipated system behaviour. Increased automation potentially decreases overall awareness of crucial system processes that are masked and leads to a degradation of manual flying skills (Sarter & Woods, 1992; Funk, Lyall, Wilson, Vint, Niemczyk, Suroteguh & Owen, 1999; Parasuraman & Manzey, 2010). Considering that pilot decision-making can be also affected by human-specific factors (e.g. cognitive biases, the influence of emotions), designers assign to aircraft automation multiple functions that are fallible to human performance imperfections. In certain instances, during adverse situations automation detects, decides and acts on behalf of the pilots, thus rendering aircraft operation mostly a procedural activity (Chialastri, 2012; FAA, 2016;

Telfer & Moore, 1997). Despite the on-going discussions about the effects of automation on human performance, Sarter & Alexander (2000) did not find significant differences in error detection and type of error between conventional and glass-cockpit aircraft.

Regarding aircraft age, Herrera & Vasigh (2009) found that the probability of an accident in commercial aviation dropped as aircraft age increased, noticed that most accidents occurred with aircraft between 15 and 29 years old, and detected a declining trend for older 'retiring' aircraft. Conversely, the work of Hansman (2014), who studied accidents between 1959 and 2012, did not support the findings of Herrera & Vasigh (2009) and did not identify significant associations between the age of commercial jet aircraft and accident rates over time. The increase of an aircraft's age has been linked to several factors that influence the sustainment of the fleet and preservation of safety standards (Colavita, Coquelet, Drury, Günther, Lincoln, Neubauer, Pfoertner, Ratwani & Sampath, 2001; ATSB, 2007). For instance, older aircraft are associated with requirements for costly updates or upgrades, increased maintenance costs, and unavailability of services and components from suppliers. Maintenance efforts focus on the structural integrity of aircraft which deteriorates over time and manifests into problems related to airframe corrosion and fatigue propagation.

Outcomes and causes of accidents have also been examined against types of operations. Evans (2007) showed that accidents in commercial operations resulted less often in a fatality compared to general aviation and commuter operations. At an international level, 76% of all accidents in the period 2012 - 2016 regarded passenger operations, while the remaining 21% and 3% of the accidents occurred during Cargo and Ferry operations respectively (IATA, 2017). In the European region (EASA, 2016a), one fatal accident, 24 non-fatal accidents and 58 serious incidents were recorded in commercial air transport in 2015, while non-commercial operations experienced 41 fatal accidents, 279 non-fatal accidents and 18 serious incidents. Regarding operation types, pilot error was a probable cause in about 40% of air transport accidents, 75% of commuter and air taxi accidents, and 85% of general aviation accidents (Li, Baker, Grabowski & Rebok, 2001; Shappell, Detwiler, Holcomb, Hackworth, Boquet & Wiegmann, 2006; Wiegmann & Shappell, 2003). Furthermore, O'Hare (2006) identified that the ability to intervene in an adverse situation was less likely in private operations than commercial ones.

The type of safety event concerned, according to Evans (2007), Loss of Control-Inflight (LOC-I), Loss of Control during Approach or Landing (LOC-A/L) and Controlled Flight into Terrain (CFIT) demonstrated the highest severities although they occurred less frequently than other event types. When aggregating the event types mentioned before, an association with fatal injuries was found at the level of 68% for commercial transport, 65% for scheduled and 55% for non-scheduled commuter operations, and 56% for general aviation (Evans, 2007). Concerning human error types, Shappell & Wiegmann (2003b) found that CFIT events did not differ significantly from non-CFIT events regarding causal factors, but the former occurrence types were associated with more violations and perceptual errors.

Moreover, various studies and reports associated the different flight phases with the occurrence of safety events and effects on human performance. In an older study using data mainly from general aviation, O'Hare, Wiggins, Batt & Morrison (1994) found that commission errors had a significant presence in the landing phase and being underrepresented in the en-route/cruise phase. Li, Baker, Grabowski & Rebok (2001) concluded that pilot error was more prevalent in airport-proximal phases (i.e. ground operations and landing/take-off) due to an increased workload compared to other flight phases. Industry reports (Airbus, 2016; Boeing, 2017) are aligned with literature and underline that about 80% of all commercial aviation accidents in the past two decades occurred during aircraft take-off, climb, descent, approach and landing.

In general, an increase in pilots' mental load during performance-demanding flight phases requires from the flight crew to prioritise attentional resources. Flight crews are able to react to a limited set of stimuli within a dynamically changing environment, thus their performance levels are lower than the ones expected or necessary to deal with an adverse situation (e.g., Baker, Qiang, Rebok & Li, 2008; Schvaneveldt, Beringer & Lamonica, 2001; Bourgeois-Bougrine, Carbon, Gounelle, Mollard & Coblentz, 2003; Tsang & Vidulich, 2006). In particular, emergency situations demand the rapid and accurate execution of actions that are time-dependent, complex and entirely situation-related. Hence, emergencies include stressful and highly variable conditions that potentially lead to severe psychological pressure on the respective controller (Bourne & Yaroush, 2003, Wise, Hopkin, & Garland, 2009). As Dismukes, Berman & Loukopoulos (2007) pointed out, human operators are ultimately susceptible to error, despite their level of expertise, and their potentiality to err is naturally increased in unfamiliar and time-dependent circumstances in conjunction with ambiguity in situational cues and perceived threat. Furthermore, O'Hare (2006) identified that the ability to intervene in an adverse situation was less likely when the operator's mental condition was adversely affected. Tightly coupled with the mental condition of the operator, fatigue remains a discernible factor contributing to degraded human performance during critical stages of the flight (e.g., Jackson & Earl, 2006; ECA, 2012).

3. Methodology

3.1 Amendment of the controllability classification

In his study, Karanikas (2015) applied the controllability classification to safety investigation reports of a single organisation, which expectedly had used a relatively similar way of report structuring and wording, this possibly affecting the reliability of the classification positively. Thus, the authors tested the reliability of the controllability classification before applying it to safety investigation reports published by different aviation investigation authorities. Four undergraduate students from the respective Faculty were briefed on the classification during a focus-group session of 90 minutes which included a presentation, an elaboration of each classification type through example cases and discussions to clarify possible misunderstandings. Afterwards, the students were asked to apply the original classification to four reports to get acquainted with the scheme and analysis process. It is noted that undergraduate students were intentionally selected due to their limited experience in aviation operations and exposure to analysis of investigation reports as a means to examine the classification's reliability regardless of the user's experience level. Under this setting, we aimed at minimising interpretation of investigation findings based on the analyst's perspectives and biases due to own experiences. The expectation was that inexperienced analysts would apply the classification by distilling information clearly stated in the reports and not extrapolating or completing it with their views.

The results of the focus-group session revealed different interpretations of the "Neutral" and "Positively Controlled" events, the definitions of which were discussed with the participants and revised accordingly. Afterwards, the five analysts (i.e. the four students and one of the authors) applied the classification individually to twenty randomly selected safety investigation reports of different length and respective event severity. Then, the data were subject to Intra-class Correlation Coefficient (ICC) calculations performed with the SPSS Version 22 software (IBM, 2013) under the settings: Two-Way Mixed, Absolute Agreement, Test Value = 0, Confidence level 95%. It is noted that ICC values range from 0, indicating a total disagreement, to 1, which reflects a complete agreement. The results showed an ICC value of 0,917 for the controllability variable and a value of 0,845 for the variable referring to the effect of the human intervention. Even though both ICC values showed a substantial agreement amongst analysts, yet different interpretations of "Neutral" and "Positively

Controlled" events affected 80% of the cases where differences were recorded. Hence, the authors decided to refine the definitions used as a means to increase further the reliability of the classification.

Taking into account that both the "Neutral" and "Controlled" event classes refer to attempts of end-users to intervene, the authors revisited the particular distinction. Instead of referring to "Neutral" as a different event category, we eliminated the "Neutral" category from the classification and we maintained only the values of "Controlled" or "Uncontrolled" in the category of controllability. Moreover, we considered the mental demands to react to an unfolding situation since, according to the literature cited above, this parameter has been seen as determinative of the success to control the course of an event as expected or prescribed (Dismukes, Berman, & Loukopoulos, 2007), a factor that has not yet been consistently examined (Sarter & Alexander, 2000; Reason, 2008). More specifically, we introduced an additional dimension of the controllability classification that refers to the system-induced type of surprise as situational or fundamental (e.g., Woods, Dekker, Cook, Johannesen & Sarter, 2010; Wiggins & Loveday, 2015). Since the difference between situational and fundamental surprises is relative to its observer (Nemeth & Hollnagel, 2016), in the context of the amended classification we defined this distinction depending on the end-user familiarity with the unfolding situation. Such a distinction can be identified in a safety investigation report based on the context of the event and the degree to which end-users employed automated skills, consulted established rules or relied heavily on their general knowledge and experience. In connection with the Skill-Rule-Knowledge taxonomy described by Rasmussen (1983) and adopted by Reason (1990) in the Generic Error-Modelling System (GEMS):

(1) Situational surprises can be managed principally through the application of prescribed procedures and/or based on training provided, requiring a combination of automated and non-fully automated skills.
(2) Fundamental surprises dictate predominantly a combination of rule-based and knowledge-based performance, without excluding skill-based reactions.

Under the approach explained above, we added the "Familiarity Level" category (Table 2). It is clarified that in the definitions of the specific category in Table 2 the reference to prescribed procedures, training, required skills and knowledge is used to reflect the match of these with the situation under study and distinguish between situational and fundamental surprises. Also, the authors contemplated that a further decomposition of the Familiarity category (e.g., low, medium and high levels separately) would render difficult to perform reliable analyses of investigation reports, because rarely there are events that include startling conditions of a single type (i.e. situational or fundamental). In the definitions provided in Table 2 for the Familiarity category, the threshold lies in the existence or not of at least one fundamental surprise.

The effectiveness of the end-users intervention in the case of a controlled event was preserved as "Positive" and "Negative" under the category named "Reaction Effect" (Table 2). Positive reaction effects are assigned to situations where the operator confronted with the situation by trying, at least, to follow respective procedures to resolve a situation and/or succeeded in the mitigation of anticipated implications, even if such resolution required actions beyond prescribed rules. The identification of the "Negative" effect applies to cases that human users' reactions led to outcomes worse than the ones expected. In this case, negative effects can illuminate potential imperfections in established procedures or training as well as local human performance problems, the whole array of organisational and environmental factors included. Overall, it is noticed that the categories of Positive and Negative effects are aimed to function as a reflection of the success or failure of the whole system and do not suggest a sole focus on individual or team performance. Inevitably, the starting point of

examining an event are the actions and accounts of end-users, but the detection of (un)successful interventions of personnel at the work floor should be viewed as indications of overall system's performance.

We would like to stress out that the controlled-uncontrolled and positive-negative categories used in the amended classification remains largely consistent with the original classification (Karanikas, 2015). The changes we introduced were based on the results of its pilot application as described above, and the interpretation of the findings of the original publication by its author in the discussion and conclusion sections of the specific paper. It is clarified that event uncontrollability does not only correspond to the unawareness of the end-user about the unfolding conditions but can also include cases that other factors deprived the operator of taking control while being aware (e.g., physical, physiology or technological limitations).

The Familiarity Level category added in the amended classification aims to serve as a bridge between the case of "controllability" (i.e. user's awareness of the unfolding adverse conditions and consequent actions to adapt to or resolve this situation, regardless of the success to completely control the event) and its effects (i.e. the extent to which the event was controlled, and its expected adverse outcomes were mitigated). Following the modification of the classification (Table 2), the five participants of the first inter-rater agreement test applied it to the same reports. All analysts classified the previously "Neutral" events as Controlled, indicating thus that the amendments substantially increased the reliability of the controllability classification. A value of 0,92 of inter-rater agreement was achieved regarding the additional column of Familiarity Level. Appendix A presents respective examples of the amended classification.

EVENT CONTROLLABILITY	FAMILIARITY LEVEL	REACTION EFFECT
CONTROLLED: The user exerted some or full degree of control over the system under responsibility before the actual (unwanted) outcomes of the event were recorded/reported.	MEDIUM-HIGH FAMILIARITY (MHF): Conditions that corresponded only to situational surprises and required the application of standard procedures described in check-lists and manuals or the implementation of resolution rules, partially or fully included in the training provided.	POSITIVE: (1) User's reactions did not worsen the expected outcome(s), or (2) regardless the outcome(s), the situation was managed properly against established procedures and/or training provided.

NEGATIVE: User's reactions following the event's initiation resulted in outcomes worse than the ones expected, whereas applicable procedures and/or training provided were seen as sufficient to deal with the situation. |
| | LOW-MEDIUM FAMILIARITY (LMF): Conditions signed by at least one fundamental surprise that dictated the combination of acquired knowledge and experience as well as existing rules/check-lists/manuals which partially (or not at all) describe respective remedial actions. | |
| UNCONTROLLED: (Unwanted) outcomes were developed without prior control of the system under concern. There had been no intervention with the scope to manage the event until the time the (adverse) outcomes were recorded. | Not applicable | Not applicable |

Table 2: Amended controllability classification of events

3.2 Application of the controllability classification

The classification of events proposed in Table 2 was applied to 296 investigation reports of aviation accidents, serious incidents and incidents. Since the classification is suggested to complement severity-based classifications by considering first an event's controllability, in our study we included events from all three severity types as a means to allow comparisons and explore associations, as explained below. The reports were available on the internet and issued by five investigation agencies and authorities, the identity of which is concealed since the objective of the research was to demonstrate the use of the classification and not to reach to conclusive results. The specific authorities/countries are hereafter labelled as Ax (i.e. x=1-5), and they were selected because they publish most or all of their safety investigations reports in English, and they sufficiently represent the so-called "Western" aviation sector.

All published and online accessible safety reports from the five authorities for events occurred between 1990 and 2014 were downloaded and administered in an Excel spreadsheet file. After that, we ran a random selection function to select approximately 40 to 60 reports per authority. Available resources and time constraints drove the sample size, and the authors recognise that the sample used in this study is not representative of the whole aviation sector or the specific States (i.e. the online repositories included hundreds to thousands of reports published in the period mentioned above). Therefore, the examination of differences across the variables, as elaborated below, were considered exploratory with the sole purpose of demonstrating the classification's discriminative power and its potential to reveal variances across several areas of interest.

Since in several instances more than one aircraft was involved in the events investigated and there had been reports referring to more than one occurrence, in total 317 cases were subject to analysis. Examples of the application of the controllability classification to the investigation reports studied are shown in Appendix B; the identity of reports is not disclosed as a means to avoid the generation of biases towards specific regions, aircraft types etc. To apply the controllability classification the authors studied all sections of the investigation reports (e.g. factual data, investigation analysis and conclusions) and recorded the values of the classification variables into a spreadsheet. The analysis of the cases necessitated the investment of approximately 320 hours. Also, the authors recorded from the reports various event characteristics and factors mentioned in the literature (see section 2.2 above) and partially included in the original study of Karanikas (2015). The event characteristics/factors (hereafter referred as event variables) were grouped based on their frequencies in the sample to achieve evenly distributed sample sizes where possible. These groups were used to conduct the statistical tests described later in this section of the paper.

Appendix C shows the event variables along with their values and distributions across the sample. The grouping categories are shown in Table 3 which also includes the classification variables against which the event variables were statistically tested for possible associations. It is clarified that:

(1) fatigue was used as variable only for the reaction effects because the particular factor is linked to human performance and not the nature of the situation or the opportunity to intervene;

(2) the event characteristics were used to explore any associations between current practice in event classification and the one suggested in this paper;

(3) the rest of the variables were used in the statistics under the concept that they represent organisational, technological and temporal factors probably linked to the events under study.

Groups of Event Variables	Classification Variables		
	Controllability	Familiarity level	Reaction effect
Origin of aircraft registration	X	X	X
Temporal and seasonal factors	X	X	X
Aircraft characteristics	X	X	X
Operational characteristics	X	X	X
Event characteristics	X	X	X
Crew fatigue	-	-	X

Table 3: Correspondence of groups of event variables with the classification variables tested

Following the calculation of the frequencies of the classification variables across the whole sample, the data, which were nominal, were subject to Chi-square tests to search for associations between the event and classification variables. The SPSS Version 22 software (IBM, 2013) was used for the statistical tests. To strengthen the validity of the results in cases of unbalanced distributions, the tests were run with the Monte Carlo simulation option of the SPSS under the default settings of Confidence Level: 99%, Number of Samples: 10.000 and starting seed 2.000.000. The significance level for all statistical tests was set to 0,05.

4. Results

The 317 cases processed were distributed across the classification characteristics as follows:

- Event controllability: 75,7% Controlled and 24,3% Uncontrolled
- Familiarity level: 32,5% Medium-High and 67,5% Low-Medium
- Reaction effect (for controlled events): 71,3% Positive and 28,7% Negative

Regarding the whole sample, flight crews positively controlled all Medium-High Familiarity (MHF) cases, whereas all negatively controlled events were associated with Low-Medium Familiarity (LMF) conditions [X^2 (1, N = 240) = 46,628, p = 0,000]. Also, 57.4% of the LMF cases were associated with positive effects and the rest 42.6% with negative effects. Table 4 presents the results of the statistical tests between the pairs of event and classification variables. For space-saving reasons and since we report the sample size of each variable in Appendix C, Table 4 includes only the degrees of freedom (DF) and the Chi-squared and p values. The significant results presented in Table 4 are reported per event variable in the subsections below. Temporal and seasonal factors were not associated with any of the classification characteristics.

Group	Variable & Degrees of Freedom	Results from Chi-square tests (significant results underlined [p < 0,05])		
		Controllability	Familiarity level	Reaction effect
Registration of the aircraft	Country (DF=5)	$\chi^2 = 11,162$ P = 0,048	$\chi^2 = 5,599$ P = 0,347	$\chi^2 = 43,352$ P = 0,000
Temporal	Year of event (DF=3)	$\chi^2 = 3,919$ P = 0,270	$\chi^2 = 1,101$ P = 0,777	$\chi^2 = 2,282$ P = 0,516
	Season (DF=3)	$\chi^2 = 1,433$ P = 0,698	$\chi^2 = 1,915$ P = 0,590	$\chi^2 = 3,288$ P = 0,349
	Daytime (DF=3)	$\chi^2 = 1,645$ P = 0,649	$\chi^2 = 2,880$ P = 0,411	$\chi^2 = 4,015$ P = 0,260
Aircraft characteristics	Age (DF=3)	$\chi^2 = 5,242$ P = 0,155	$\chi^2 = 8,200$ P = 0,042	$\chi^2 = 3,752$ P = 0,290

Group	Variable & Degrees of Freedom	Results from Chi-square tests (significant results underlined [p < 0,05])		
		Controllability	Familiarity level	Reaction effect
	Type (DF=2)	χ^2 = 17,028 P = 0,000	χ^2 = 10,732 P = 0,005	χ^2 = 2,805 P = 0,246
	Weight (DF=1)	χ^2 = 19,993 P = 0,000	χ^2 = 7,998 P = 0,005	χ^2 = 2,345 P = 0,126
Operational characteristics	Type (DF=1)	χ^2 = 19,249 P = 0,000	χ^2 = 4,937 P = 0,026	χ^2 = 0,806 P = 0,369
	Scope (DF=1)	χ^2 = 10,670 P = 0,001	χ^2 = 7,109 P = 0,008	χ^2 = 2,762 P = 0,097
	Phase (DF=2)	χ^2 = 0,313 P = 0,855	χ^2 = 12,132 P = 0,002	χ^2 = 7,474 P = 0,024
Event characteristics	Type (DF=6)	χ^2 = 39,515 P = 0,000	χ^2 = 32,651 P = 0,000	χ^2 = 32,402 P = 0,000
	Severity (DF=2)	χ^2 = 12,913 P = 0,002	χ^2 = 26,678 P = 0,000	χ^2 = 34,360 P = 0,000
Crew/ground staff fatigue	Fatigue (DF=1)	N/A*	N/A*	χ^2 = 18,716 P = 0,000
* N/A: Not Applicable				

Table 4: Statistical results of the associations between classification and event variables from Chi-square analysis.

4.1 Registration of the aircraft

As shown in Figure 1, when considering the whole sample, the safety events of the flight operators registered in A2 were controlled to a 91,7% level, where the events of operators from A5 showed the minimum controllability with 63,6% controlled events. The controlled events concerned (Figure 1), interventions of flight crews from A3 and A2 operators resulted to 95,2% and 90,9% positive effects respectively, whereas the ones of the operators registered in A4 demonstrated the least positive effects at a level of 44,6%.
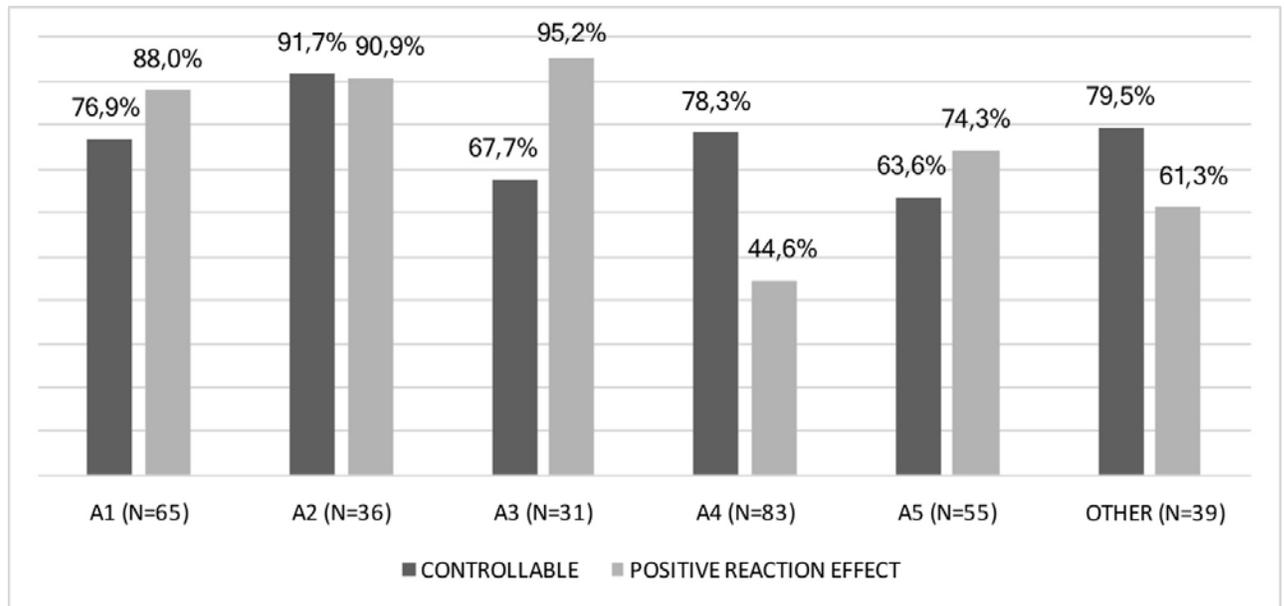
Figure 1: Frequency distribution per analysed country of registration for controlled events and positive reaction effects.

4.2 Aircraft characteristics

Most of the Low-Medium Familiarity (LMF) levels were found in operations with aircraft in the highest age cluster (i.e. over 25 years, 77,6%), whereas newest aircraft were associated with the highest frequency of Medium-High Familiarity (MHF) events (i.e. ≤ 6 years, 45,3%). Jet aircraft were involved in more controlled events (84,9%) compared to the propeller (70,7%) and rotary aircraft (56,8%). The respective LMF/MHF percentages were 58,9%/41,1% for jet, 76,8%/23,2% for propeller and 84,0%/16,0% for rotary type aircraft.

Regarding the aircraft weight, heavier aircraft were associated with more controlled occurrences (88,1%) compared to aircraft weighing less than 27 tons (i.e. 66,1% controlled events). Aircraft weighing more than 27 tons were involved in LMF/MHF cases with frequencies of 59,3%/40,7%, whereas the corresponding frequencies for lighter aircraft were 76,5%/23,5%.

4.3 Flight characteristics

Events during CAT operations were controlled in 81,5% of the cases, whereas events of other operation types were controlled at a level of 57,1%. CAT type events were associated with 33,1% MHF and 66,9% LMF conditions, and the respective frequencies for other operation types were 16,7% for MHF and 83,3% for LMF.

Passenger flight events were controlled in 82,3% of the cases, and non-passenger flight events were controlled at a level of 65,9%. Passenger flight events required reactions to LMF/MHF situations in a 63,7%/36,3% ratio and non-passenger flights were found with 80,2% LMF and 19,8% MHF events.

The ratio of LMF/MHF cases was 77,1%/22,9% for en-route events, 72,0%/28,0% for other flight phases and 51,4%/48,6% for events occurred during ground operations. Ground events were associated with more positive intervention effects (i.e. 80,0%) compared to cases with events en-route (i.e. 75,7%) or during other flight phases (i.e. 62,0%).

4.4 Event characteristics

The distribution of controllability, reaction level and effects per event category are presented in Figure 2 where numbers have been rounded to integers to improve readability. Runway excursions (RE) and technical failures [Power Failures (PF) and Other Technical Failures (OTF)] were mostly controlled (i.e. about 92% to 96% of the cases) whereas Controlled Flight into Terrain (CFIT) events were the least controlled ones with about a 50% (un)controllability. Mid-air collisions (MAC) were the cases with the fewest LMF conditions present (i.e. 26%), and PF and Loss of Control In-flight (LOC-I) occurrences were connected with LMF conditions at a level between 80% and 90%. The CFIT cases were mostly a result of negative control effects at a level of about 64%, and MAC events involved the least negative effects in 9% of the cases.
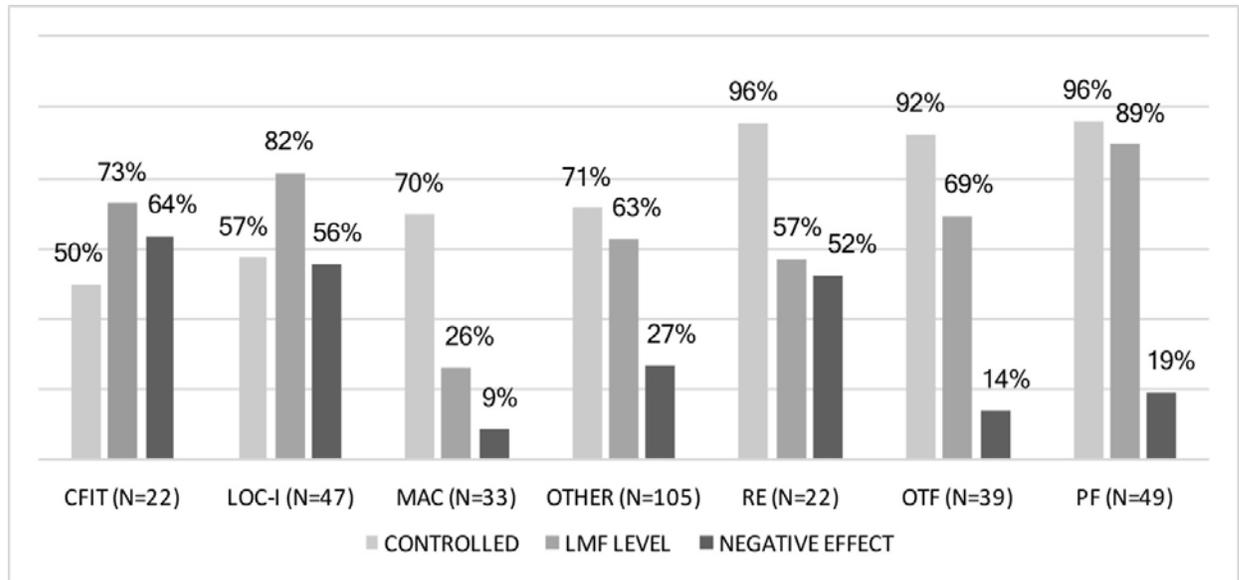
Figure 2: Frequency distribution per event category for controlled events, LMF levels and negative reaction effects.

Decreasing controllability, decreasing MHF conditions and increasing negative effects of intervention corresponded to the ascending order of the events' severity (Figure 3):

- Incidents: 93,8% controllable / 55,6% MHF / 6,7% negative effects
- Serious incidents: 80,3% controllable / 47,2% MHF / 9,4% negative effects
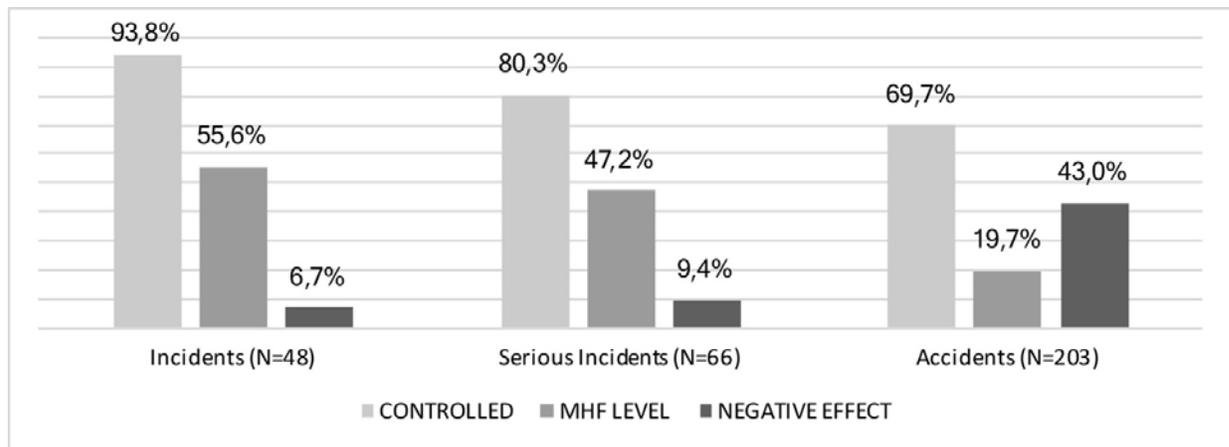- Accidents: 69,7% controllable / 19,7% MHF / 43,0% negative effects

Figure 3: Frequency distribution per event severity category for controlled events, MHF level and negative reaction effect.

4.5 Fatigue

Where fatigue was recorded as a contributing/causal factor to the event, increased negative effects (i.e. 66,7%) were found in contrast to the cases that fatigue was not mentioned in the investigation report as a factor (i.e. 75,5% positive effects).

5. Discussion

5.1 Overall picture

The results suggest that about three-quarters of the events studied were controlled by the end-user(s) who in those cases had the opportunity to mitigate an event's outcome. However, about a quarter of the outcomes which dictated an investigation was either noticed after the window of effective action or not noticed at all during the flight, meaning that several (serious) incidents could have escalated into accidents. This means that about 6% of the incidents and 20% of the serious incidents included in this research were uncontrolled and could have resulted in accident-level consequences. Hence, on the one hand, taking into account that the number of incidents investigated is only a small portion of the ones recorded, it seems that aviation might have missed critical information about the degree to which its systems are vulnerable, and their outcomes are controlled. This finding confirms the concerns of Wise, Hopkin & Garland (2009) and ICAO (1993) who claimed that incidents might include valuable material to be used for improving safety.

As the results about uncontrollability suggest, the ever-decreasing accident rates could have been higher, thus leading to past safety performance levels different than the ones communicated nowadays. Somewhat surprisingly, the controllability of events did not change significantly over time, albeit published industry reports show a declining trend of accident rates. Moreover, it could be claimed that any hypothetical occurrence rates based on the potential of (serious) incidents to escalate into accidents would be different if uncontrolled eventualities recorded through voluntary reporting systems and not subject to investigations had been considered.

Notwithstanding, more than three-thirds of the controlled events necessitated reactions to cope with Low-Medium Familiarity (LMF) cases, indicating that the confrontation with most of the situations that resulted in investigated cases was not adequately prescribed in standard procedures and/or sufficiently supported by pilot training. Notably, though, most controlled situations resulted in alleviated end-states as indicated by the high percentage of positive reaction effects (71,3%), meaning

remarkable levels of preparedness and responsiveness of the persons actively involved in the events. Thus, when the pilots had the opportunity to intervene in the course of events, this most of the times led to successful mitigation of the possible negative impacts. Interestingly, negative effects of interventions were recorded only in LMF cases, whereas all Medium-High Familiarity (MHF) events were associated only with positive intervention effects. The former finding is aligned with the literature discussing the difficulty of end-users to timely resolve unfamiliar situations under stressful conditions (Bourne & Yaroush, 2003; Dismukes, Berman & Loukopoulos, 2007; Wise, Hopkin, & Garland, 2009). The association of MHF cases with positive effects indicates the merit of training and standard procedures when those apply to the situations encountered (Dismukes, Berman & Loukopoulos, 2007; Telfer & Moore, 1997). It is clarified that the authors do not suggest that the industry shall attempt to prescribe procedures for all situations and prolong pilot training. Such attempts would be unfeasible due to the vast number of possible combinations of factors that can lead to unforeseeable system states. Nevertheless, we should not overlook the value of consistent, focused and effective training within the limits of given resources and the boundaries of human performance.

5.2 Event controllability and severity

The severity bias included in the sample of the current study led expectedly to a significant association of the severity class with the outcome effectiveness frequencies; the more the negative effects of human intervention, the higher the severity. However, despite this expected link, the results revealed that in 57% of the accidents and about 91% of the serious incidents, which are the types of events that are obligatorily investigated, the flight crews controlled the situations by successfully mitigating the negative outcomes of the events. This finding indicates areas for positive learning that under the dominant outcome-focused approach might not have been fully exploited during investigations. Therefore, an examination of the successes, as proposed by the Safety-II approach (Hollnagel, 2014), does not necessarily bring a need to investigate more safety events under the reality of limited resources. Although incidents might remain a source of valuable information to be distilled during selected investigations (Wise, Hopkin, & Garland, 2009), serious incidents and accidents can yet comprise opportunities to collect information about how end-users managed to deal effectively with adverse, unexpected conditions.

Regarding the familiarity levels, the findings showed that the accidents were associated almost exclusively with Low-Medium Familiarity (LMF) conditions. This confirms that extremely negative outcomes are coupled with the requirement to collect, process and use information to build an adequate situational awareness of an on-going, partially or wholly unfamiliar, event, and act respectively, usually under time and space constraints (Wise, Hopkin & Garland, 2009; Dismukes, Berman & Loukopoulos, 2007). Nonetheless, the considerable number of uncontrolled and LMF types of incidents and serious incidents should be noted, also taking into account that incidents are very likely to go unreported if their outcomes are within the expected margins or they are not visible and announced outside the cockpit environment (e.g. Bhagwati, 2006).

5.3 Differences across regions

From a regional and comparative perspective, it seems that A2-registered crews outperformed in both the controllability of the events and the frequency of positive effects of interventions, whereas the picture varied and was occasionally controversial across the rest of the registration countries. For example, A3 operators were involved in controlled events less frequently than most of the other countries included in this study, and they presented the highest frequency of positive effects. However, events with A4-registered aircraft had a higher controllability frequency than the A3 but

demonstrated the lowest percentage of positive effects of its crew amongst the regions included in this research.

Although the frequencies per controllability variable are only indicative and limited to the sample studied, the findings suggest that there might be large controllability-related differences between regions even within the Western aviation sector. Such differences can be attributed to the various degrees of inclusiveness and effectiveness of crew training and different levels of pilot competency in managing adverse situations, such factors, in turn, reflecting deeper and broader issues. Also, it is possible that negatively influenced events are reported at different extents across different regions, this possibly reflecting cultural differences, especially ones connected with the just and reporting types of safety culture (Reason, 1998). The authors would like to stress out that this regional analysis was not performed to raise concerns about specific geographical areas, rather to demonstrate the classification's ability to reveal other aspects of interest when comparing multiple locations.

When the authors considered the accident rates of years 2008-2016 (ICAO, 2017b) and the controllability ratio of incidents and serious incidents combined for each country for the same period, it seems that reliance on accident rates might have concealed important messages if the controllability of occurrences less severe than accidents is not contemplated. Country A3, for instance, achieved seven accident-free years, but more than a quarter of its (serious) incidents included in this study were uncontrolled and could have escalated into accidents. On the other hand, events with low and medium severities for A4-registered aircraft showed full controllability, but, contradictory to A3, the A4 country recorded an overall higher accident rate and no accident-free year. Therefore, A4 accident rates could represent its past safety performance more reliably since its less severe eventualities were controlled entirely. It is reminded that the controllability variables were not statistically altered over time (see section 4 above). Hence, the respective figures of accident rates and controllability classification results for the period above are claimed as comparable although limited to the reports studied in this research.

The example given above indicates that an emphasis on accident rates alone might generate complacency and mask valuable messages that the contemplation of event controllability might convey. Demonstrating safety performance merely through actual outcomes of occurrences might deprive States and organisations of examining events that had not escalated into accidents due to chance. It is clarified that the authors do not discuss the accident rates against the reaction effects because the latter do not characterise an event's potential elevation into a higher severity class but indicates the results of end-users attempt to cope with an adverse situation. For instance, a positive reaction could have significantly reduced the number of casualties during an occurrence, yet a single fatality would be sufficient to classify the specific occurrence as an accident.

5.4 Differences across other event variables

The aircraft characteristics concerned, the results showed that younger, heavier and jet-type aircraft required from flight crews fewer cognitive resources to cope with unfolding adverse situations (i.e., more MHF levels), while the last two categories mentioned were also involved more often into controlled events. These findings align with industry reports (e.g. Airbus, 2017) stating that technological advancements implemented in new generation aircraft have improved aviation safety due to the introduction of automated control systems (e.g. Ground Proximity Warning System). Such systems aim at optimising the quality and quantity of information flow to pilots and provide support to crews to maintain operational safety limits. Automation and event-based training are sought to shift the nature of unexpected conditions from fundamental surprises to situational ones (FAA, 2016; Chialastri, 2012) as a means to minimise resource demands.

Also, the differences regarding the controllability of events and the type of responses required were similarly significant between CAT and non-CAT operations and between passenger and other types of flights. CAT and passenger operations were more controllable and required reactions to more MHF cases than necessitated by non-CAT and non-passenger ones, as also shown in the work of O'Hare (2006). These differences possibly reflect the type and age of aircraft used along with the respective level of automation. CAT and passenger flights are typically operated with younger, larger and jet aircraft compared to non-CAT and non-passenger flights (IATA, 2015; GAMA, 2017).

However, various authors have claimed that the integration of highly automated systems in modern aircraft has deteriorated manual flying skills and adversely affects pilots' situational awareness (e.g. Sarter & Woods, 1992; Funk et al., 1999; Parasuraman & Manzey, 2010). Therefore, when considering the findings of this work in conjunction with the suggestions of the literature regarding the effects of automation, it seems that a delicate balance between the level of flight automation and the necessity to maintain control and awareness of system processes is necessary. On the one hand, according to the results of this study, automation seemed supportive in minimising fundamental surprises in the cockpit and rendering eventualities more controllable. On the other hand, pilots must be still capable of dealing with unprecedented and unfamiliar conditions sourcing from design flaws and/or any type of excessive disturbance, the effects of which have not been anticipated and their resolutions have not been embedded in automated functions.

Regarding the flight phase, adverse situations encountered in en-route, climb or descend flight phases required more actions related to LMF conditions and recorded more negative reaction effects than the cases of ground events. This finding might be attributed to an increased standardisation and automation support of ground movements due to an association of the particular flight phase with increased workload in earlier studies (e.g. Li, Baker, Grabowski & Rebok, 2001). Furthermore, the specific results can be linked to the relatively increased available time to respond to ground events which allows the pilots to recall procedures, apply best practices learned from training, cross-check information etc. The association of events on the ground with more MHF cases compared to other flight phases aligns with the findings of O'Hare, Wiggins, Batt & Morrison (1994) who found that simple errors in procedures' execution were most prevalent in the ground operations segment.

Runway excursions and technical failures showed remarkable controllability levels, but half of the former cases were associated with negative intervention effects, whereas the reactions to technical failures were mostly successful. The often-negative results of runway excursions have been a concern of the industry and the findings of this study accord with the initiatives taken to decrease those types of events and improve the success rates when dealing with those (ICAO, 2013b, 2017a). The high frequency of successes when experiencing technical failures might be explained by the fact that resolutions of all foreseen technical problems are described in flight manuals (e.g. Airplane Flight Manual, Quick Reference Handbook, or Operations Manual) or displays and practised in initial and recurrent training as well as during periodical proficiency checks. However, the results interestingly revealed that the majority of the technical failures were Low-Medium Familiarity cases. This finding can be attributed to the fact that, instead of a single technical failure, on several occasions system errors co-occur or happen in rapid succession, hence generating complex circumstances (Maurino, Reason, Johnston & Lee, 2017; Woods, Johannesen Cook, & Sarter, 1994; Plioutsias & Karanikas, 2015).

Nevertheless, the findings of this study suggest that crews in most of the cases with technical problems were armed with the skills, knowledge and experience to face those and avoid or mitigate negative safety implications. Hence, future research could examine whether the structure, content and quality of the programs and practices followed for dealing with technical failures can be customised

appropriately and transferred to the training for other types of events that, according to this study, seem more vulnerable (i.e. runway excursions, loss-of-control in flight and controlled-flight-into-terrain events). Also, any future initiative must not underestimate the effects of fatigue inflicted by individual habits, local working environment issues, ineffective human-machine interaction or other factors. In this study, contradictory to the results of O'Hare (2006), degraded human performance, as demonstrated through an unsuccessful intervention into the course of the event, was associated with cases that fatigue was stated in the investigation reports as a causal or contributing factor.

Markedly, the positive or negative effects of pilots' reaction were not statistically associated with the aircraft characteristics considered in this study and indirectly reflecting factors related to the human-machine interaction. Thus, although aircraft characteristics probably played a role in the event controllability and level of familiarity, the success of the crew interventions lied only on factors linked to individual/team characteristics, those varying across the geographical areas included in this research and presumably reflecting effects of latent and wider organisational and regional factors.

5.5 Contextualisation of the results

When comparing the findings of this study with the results of the application of the original classification to a single organisation (Karanikas, 2015), it seems that in the latter there were more uncontrolled events. This difference can be attributed to the fact that the specific organisation was investigating internally all low-severity events and not only occurrences with highly adverse outcomes, which for a single organisation are rarer. This extended the sample studied in the research of (Karanikas, 2015) and led to a quite different distribution of event severities between his study and the current research. However, this observation might denote that the average controllability of all incidents might be lower than the one calculated for investigated incidents which, according to the current practice, are a small fraction of the ones occurring.

The human intervention concerned, a high percentage of positive reaction effects and invariance of controllability frequency over time were detected for the events of the organisation studied by (Karanikas, 2015) similarly to the results of this research. Also, expectedly and analogously to the results of the current study, the low-severity occurrences of the single organisation were associated with increased controllability compared to the high-severity ones. Interestingly, although the controllability and reaction effects in the work of Karanikas (2015) were associated with the age of aircraft, this was not revealed from this study. The authors believe that this difference is attributable to the smaller diversity of aircraft fleet operated by a single organisation compared to the various types of aircraft involved in the events included in this research.

5.6 Controllability classification and contemporary safety thinking

The controllability classification suggested in this paper is aligned with the thinking that underlies Resilience Engineering (Hollnagel, Woods, & Leveson, 2006) since we advocate the examination of whether the operators could deal with situational or fundamental surprises that correspond to less or more unforeseen or unexpected situations. Actions to recover or alleviate safety-critical situations by the operator in such complex systems sometimes require actions that surpass existing rules and prescribed procedures (Bergström, et al., 2009). Under this concept, the results of our study have shown that situations quite familiar to the operator were always successfully managed and matched the training provided as well as prescribed procedures and/or rules, hence showing the importance of effective training and balanced guidelines. Regarding Low-Medium Familiarity conditions with positive reaction effects, the results of the current study highlighted events that required reactions not (fully) anticipated for the respective system (i.e. not matching with available expertise, training, procedures and rules or unforeseen excessive disturbances). The specific cases indicated resilient systems that

allowed operators to react to a given situation and alleviate, to the extent possible, the outcome of an event.

Therefore, the application of the suggested classification can lead to the detection of either adequately resilient systems, hence creating an opportunity to learn from successes, or systems that may require additional efforts to build the level of resilience that will enable them to confront with situations falling outside the boundaries of procedures-based management. Even further, taking into account the commonly accepted cornerstones of resilience (Patriarca, Bergström, Di Gravio, & Costantino, 2018), the controllability classification can function as a monitoring mechanism of the extent to which a system provides the opportunity to its end-users to respond (i.e. Controllability variable), anticipates excessive disturbances and variability (i.e. Familiarity Level), and has learned from past successes and failures (i.e. Reaction effects). Certainly, the controllability characteristics do not reflect the overall system resilience, but the authors contemplate that they can be useful as a starting point to materialise the basic resilience functions.

Moreover, the thinking behind the controllability classification can be useful in the conduction of safety investigations as a means to look with a forward time-line and minimise hindsight bias and an overemphasis on human error (Dekker, 2006). Investigators can first search for factors related to the degree to which the end-user had the opportunity to control the situation, then the extent to which the situation was (un)familiar along with corresponding explanations, afterwards the effects of intervention against expectations or prescribed rules, then the suitability and adequacy of training and procedures for the given conditions, and, lastly, possible individual or team performance issues. Under this approach, although mistakes are inherently connected to the human nature (Reason, 2008), the examination of human performance degradation and the revelation of related causes, a practice that stands predominantly central in regulatory activities, organisational safety management and investigations (Hollnagel, 2014), will constitute the end and not the starting point of investigations.

5.7 Opportunities for further developments

The uncontrolled events were not categorised because the concept of controllability as characteristic of an event had not been considered by investigators. Thus the reasons for uncontrollability were only occasionally addressed in safety investigation reports. Consequently, the authors would not reliably retrieve relevant information from the sample included in this work even if they had suggested uncontrollability factors. Nonetheless, various reasons might lie behind uncontrolled events during which the end-user was not aware of the unfolding situation or not capable of dealing with it.

The authors contemplate that the System-Theoretic Accident Model and Processes (STAMP) introduced by Leveson (2011) and the Functional Resonance Analysis Method (FRAM) suggested by Hollnagel (2012) can comprise the basis for classifying uncontrolled events. According to the former author, adverse events occur principally due to uncontrolled interactions between system parts and can be attributed to flawed control algorithms (e.g., missing procedures, lack of knowledge about system behaviours) and imperfect mental models (i.e. inadequate or delayed information flow from other system controllers or within the system). Such factors along with the rest of the keywords used to reveal causal factors for ineffective end-user actions (Leveson, 2011) might be used by human factors analysts to explore reasons for lack of system control and human underperformance (Karanikas, 2017). Based on the work of Hollnagel (2012), uncontrolled systems could be the result of mismanaged variability within the wider system due to missing, mismatching or out-of-range parameters that are essential to successful system performance (i.e. inputs, outputs, time, control, resources and preconditions).

5.7 Limitations

Apart from the limited sample used, it may be argued that the findings of this research might have been affected due to the change of investigation practices during the study's timeframe of 25 years (e.g. different investigation and analyses techniques, available human and technical resources and time invested per investigation). Similarly, variances between the different investigation agencies included in this study might have existed regarding how investigations are executed and reported (e.g. length and depth of investigations or reports), prevailing legislation per region, activity volumes (e.g. more or less General Aviation traffic in various countries) etc. However, since the classification of controllability refers to the core development of a safety-critical situation, we believe that changes in investigation practices might have only a little effect. The authors were able to record the values of the controllability classification for all events, regardless depth, length, region, publication year etc. of the investigation reports.

Also, the ambiguities in the threshold between incidents and serious incidents currently used in aviation might have affected the validity of the associations investigated in this study. We believe that the execution of future studies with larger and more representative samples could compensate for the variability in the interpretation of the lines between the severity categories used in the aviation sector. The time necessary to apply the classification to investigation reports may vary substantially due to various factors (e.g. length and/or depth of reports, reading speed driven by relevant experience), but the 320 hours invested in this study for the analysis and cross-check of the results of 317 cases can comprise a good indication of the resources required.

Moreover, we recognise that the users of the classification must become familiar with its definitions, and sufficient inter-rater agreement amongst analysts must be achieved to ensure the reliability of any study. An important point is that the controllability classification applies to a specific system and well-defined system objectives because the uncontrolled event and outcomes of one subsystem (e.g., aircraft operation resulting in a runway excursion) can be successfully dealt by another subsystem (e.g., emergency services). It is also noted that the attribution of the familiarity level as Low-Medium or Medium-High depends on the interpretation of operational complexity and the context of each situation.

6. Conclusions

6.1 Messages from retrospective analysis

This paper refined the controllability classification introduced originally by Karanikas (2015), enriched it with the characteristic of end-users' familiarity level (i.e. Low-Medium or Medium-Low), improved its reliability, and applied the amended classification to 317 cases investigated by the agencies of five countries. The results of the analysis showed that a quarter of the outcomes of the investigated events included in the sample were uncontrolled by the end-users, three-quarters of the controlled occurrences studied demanded increased cognition levels to confront with fundamental surprises, and most of the controlled situations were successfully dealt by the end-users regarding the mitigation of their expected negative outcomes. Also, the findings confirmed that cases with lower controllability of events and lower familiarity were associated with higher levels of actual severity.

From a regional perspective, the study showed significant differences amongst countries regarding the level of controllability and the degree of positive effects from users' intervention into unfolding events. Although these results are not aimed at benchmarking regions due to the limited sample used per investigation authority, the findings of the study suggest that a sole focus on accident rates might distort the picture of safety performance in overall and per geographical area. The aircraft

characteristics concerned, the results suggested that older, smaller and non-jet aircraft, which are the ones mainly outside the scope of automation sophistication, offered decreased controllability and were involved in events with Low-Medium levels of end-user familiarity. Events during flight and occurrence types other than runway excursions and technical failures showed decreased controllability characteristics, and the latter occurrence types were associated with more fundamental surprises. The particular findings indicate areas of improvement in the aviation industry to increase the controllability of events, minimise startles and increase positive intervention effects.

Furthermore, the results of this research suggest that contradictory to the decrease of accident rates over the last decades, the frequencies of controllability, familiarity levels and reaction effects have not significantly changed over time. This denotes that the aviation industry in general, albeit the developments regarding technology, training, standardisation etc., might not have adequately managed to increase the level to which the outcomes of occurrences are not a matter of luck, decrease the requirement for crews to invest increased cognitive resources to cope with unfolding, unfamiliar events, and support end-users in their attempts to mitigate negative results of controlled events. Despite the large number of reports analysed, the relatively small sample of investigated events analysed when compared to the whole available set and the grouping of variables adopted in this study does not allow the generalisation of the results of this research to the whole aviation sector. However, regardless of the limited value of the frequencies of the various controllability values, we contemplate that the findings of the statistical associations presented in this paper could be of interest and the results in overall can trigger a focus on less outcome-biased approaches to occurrence classification by considering the characteristics of the classification proposed.

6.2 Further practical value

Although in this research the controllability classification was used retrospectively to analyse past events and is inevitably subject to hindsight bias, in addition to the mandatory investigation of serious incidents and accidents the specific classification can comprise a prioritisation criterion to select which incidents can be investigated under the reality of limited resources. An initial evaluation of incidents regarding their controllability can support decision-making for proceeding with their investigation. Taking into account the high percentage of positive reaction effects during controlled events, the authors contemplate that it is more important to examine with higher priority incidents whose outcomes were a matter of chance and not a result of end-users intervention. Even more, the order to which the classification variables are presented can be useful during investigations. A focus on whether the end-user was provided with the opportunity to intervene into an event or affect its outcomes could be followed by the degree to which the conditions of the event were (un)familiar and then whether there was a match of the procedures and personnel skills, knowledge and experience with the situation under study.

Moreover, the characteristics of the controllability classification could be added in voluntary reporting systems, where the end-users could state their perceived degree of controllability and level of familiarity through the introduction of suitable measurement scales, the design of which was outside the scope of this study. Certainly, the extent to which a situation is familiar to the operator and the cognitive and physical capacity necessary to cope vary per person due to different experiences, various emotional, psychological and physical states and diverse sets of skills. Thus, such measurements could be based on a Likert-scale concept where operators would rate the degree of control over their system during a particular situation, the extent of familiarity with the situation, the degree to which procedures and training were sufficient and matched the conditions experienced, and the extent to which the actual outcomes were close to the ones anticipated when trying to control the event's march. Indeed, apart from the extreme points of any scale that might be interpreted consistently (e.g.,

"not at all" and "completely"), this type of self-assessment is expected to be quite subjective. However, on the one hand, we cannot neglect that the assessment of the conditions by the end-users would allow the collection of valuable information about their perspectives, and, on the other hand, any data could be analysed per homogeneous group to achieve comparability (e.g., experience levels, regions, operation types). In such a way, the rating of the controllability characteristics would not be binary, as they inevitably are when analysing investigation reports, and a more representative picture could be obtained about normal, daily operations. Also, risk assessments based on (un)controllability of occurrences could target the most affected areas of operations.

6.3 Final remarks

In support of modern approaches to safety that suggest more efforts in learning from successes (Safety-II) in addition to failures (Safety-I), the controllability classification distinguishes situations with and without human-system interaction, separates procedural reactions from complex cognitive ones, and displays whether control attempts were poorly or well executed against predetermined operating procedures. Regarding especially the reaction effectiveness, the authors stress out that any results should not be used to blame involved parties. The findings from any application of the suggested classification must aim at providing insights into areas that necessitate more attention and require further studies to offer end-users the opportunity to intervene into adverse system states and support them in their honest attempts to successfully manage adverse situations, possibly based on the aftermaths of adequately controlled events.

Overall, the analysis of the data included in this paper demonstrated that the proposed classification provides an alternative, less outcome-centric approach to safety events. It might be used by any industry sector as a more valid safety performance indicator compared to the mere use of rates of occurrences per severity level. The high reliability achieved is also promising for benchmarking different sectors and establishing a common approach to safety events. The inclusion of the particular classification in voluntary safety reporting systems could serve as metric of system resilience, its use in mandatory safety reporting systems can support the prioritisation and direction of safety investigations, and its application retrospectively on investigated events could deliver important messages for past performance and aftermaths from both successes and failures. Further research is suggested to define characteristics of uncontrollability as a means to provide further insights into factors that can deprive the users of controlling their systems.

References

Airbus. (2016). *A Statistical Analysis of Commercial Aviation Accidents 1958-2015.* Blagnac Cedex, France: Airbus S.A.S.

Airbus. (2017). *A Statistical Analysis of Commercial Aviation Accidents 1958-2016.* Blagnac Cedex, France: Airbus S.A.S.

Allianz. (2014). *Global Aviation Safety Study: A review of 60 years of improvement in aviation safety.* Munich: Allianz Global Corporate & Specialty SE.

ATSB. (2007). *How Old is Too Old? The impact of ageing aircraft on aviation safety. Aviation Research and Analysis Report –B20050205.* Canberra, Australia: Australian Transport Safety Bureau.

Baker, S. P., Qiang, Y., Rebok, G. W., & Li, G. (2008). Pilot error in air carrier mishaps: longitudinal trends among 558 reports, 1983-2002. *Aviation, Space, and Environmental Medicine, 79*(1), 2-6. doi:10.3357/asem.2200.2008

Bellamy, L. J., & Sol, V. M. (2012). *A literature review on safety performance indicators supporting the control of major hazards. RIVM rapport 620089001.* Bilthoven, the Netherlands: National Institute for Public Health and the Environment.

Bergström, J., Dahlström, N., Van Winsen, R., Lützhöft, M., Nyce, J., & Dekker, S. (2009). Rule-and role-retreat: An empirical study of procedures and resilience. *Journal of Maritime Research, 6*(3), 41-58.

Bhagwati, K. (2006). *Managing Safety.* NJ: Wiley – VCH.

Boeing. (2017). *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-2016.* Seattle: Boeing Commercial Airplanes.

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblentz, A. (2003). Perceived fatigue for short-and long-haul flights: a survey of 739 airline pilots. *Aviation, Space, and Environmental Medicine, 74*(10), 1072-1077.

Bourne, L. E., & Yaroush, R. A. (2003). *Stress and cognition: A cognitive psychological perspective. NASA/CR-2003-212282.* California: Ames Research Center.

Chialastri, A. (2012). Automation in aviation. In F. Kongoli, *Automation* (pp. 79-102). InTech.

Colavita, M., Coquelet, M., Drury, C., Günther, G., Lincoln, J., Neubauer, M., . . . Sampath, S. (2001). *Aging Aircraft Fleets: Structural and Other Subsystem Aspects. RTO-EN-015 AC/ 323(AVT-053)TP/ 33. .* Neuilly-Sur-Seine Cedex, France: North Atlantic Treaty Organization, Research and technology organization.

Dekker, S. (2006). *The field guide to understanding human error* (2nd ed.). Aldershot: Ashgate.

Detwiler, C., Hackworth, C., Holcomb, K., Boquet, A., Pfleiderer, E., Wiegmann, D., & Shappell, S. (2006). *Beneath the tip of the iceberg: A human factors analysis of general aviation accidents in Alaska versus the rest of the United States. DOT/FAA/AM-06/7.* Washington: Office of Aerospace Medicine, Federal Aviation Administration.

Di Gravio, G., Mancini, M., Patriarca, R., & Costantino, F. (2015). Overall safety performance of Air Traffic Management system: Forecasting and monitoring. *Safety Science, 72*, 351-362. doi:10.1016/j.ssci.2014.10.003

Dismukes, R. K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents.* Aldershot: Ashgate Publishing, Ltd.

Duijm, N. J. (2015). Recommendations on the use and design of risk matrices. *Safety Science, 76*, 21-31. doi:10.1016/j.ssci.2015.02.014

EASA. (2016a). *Annual Safety Review.* Cologne: European Aviation Safety Agency.

EASA. (2016b). European Plan for Aviation Safety 2016-2020. Cologne: European Aviation Safety Agency.

EC. (1996). Guidance on risk assessment at work. Luxembourg: European Commission: Office for Official Publications of the European Communities.

ECA. (2012). *Barometer on pilot fatigue.* Brussels: European Cockpit Association.

EMSA. (2016). Annual Overview of Marine Casualties and Incidents 2016. European Maritime Safety Agency.

EU. (2009). DIRECTIVE 2009/18/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 April 2009 establishing the fundamental principles governing the investigation of accidents in the maritime transport sector . European Union.

EU. (2010, October). Investigation and Prevention of Accidents and Incidents in Civil Aviation and Repealing Directive 94/56/EC. *REGULATION (EU) 996/2010*. European Union.

EU. (2016). DIRECTIVE 2016/798 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 May 2016 on railway safety. *Official Journal of the European Union, L. 138*, 102–149.

EUAR. (2016). *Railway Safety Performance in the European Union.* Valenciennes: European Union Agency for Railways.

Eurocontrol. (2009a, December 2). Reporting and Assessment of Safety Occurrences in ATM. *ESARR 2, 3rd*. Brussels: European Organisation for the Safety of Air Navigation.

Eurocontrol. (2009b, November). The Aerospace Performance Factor (APF). Developing the EUROCONTROL ESARR 2 APF. Brussels: European Organisation for the Safety of Air Navigation.

Evans, J. K. (2007). *An Application of CICCT Accident Categories to Aviation Accidents in 1988-2004. NASA/CR-2007-214888.* Hanover, MD: US National Aeronautics and Space Administration.

FAA. (2016). *Pilot's Handbook of Aeronautical Knowledge, FAA-H-8083-25B.* Washington: U.S. Department of Transportation: Federal Aviation Administration.

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight deck automation issues. *The International Journal of Aviation Psychology, 9*(2), 109-123. doi:10.1207/s15327108ijap0902_2

GAMA. (2017). *2016 General Aviation Statistical Databook & 2017 Industry Outlook.* General Aviation Manufacturers Association.

Gaur, D. (2005). Human factors analysis and classification system applied to civil aircraft accidents in India. *Aviation, Space, and Environmental Medicine, 76*(5), 501-505.

Greenwell, W. S. (2003). Learning lessons from accidents and incidents involving safety-critical software systems. Doctoral dissertation, University of Virginia.

Hansman, J. R. (2014). *Analysis of Impact of Aircraft Age on Safety for Air Transport Jet Airplanes.* Cambridge, MA: MIT International Center for Air Transportation (ICAT), Massachusetts Institute of Technology.

Herrera, J. M., & Vasigh, B. (2009). A basic analysis of ageing aircraft, region of the world, and accidents. *Journal of Business & Economics Research, 7*(5), 121-132. doi:10.19030/jber.v7i5.2299

Hollnagel, E. (2012). *FRAM – The Functional Resonance Analysis Method.* UK: Ashgate.

Hollnagel, E. (2014). *Safety-I and Safety-II: The Past and Future of Safety Management.* Farnham, UK: Ashgate.

Hollnagel, E., Woods, D. D., & Leveson, N. C. (2006). *Resilience engineering: Concepts and precepts.* Aldershot: Ashgate.

HSE. (2016). *Statistics on fatal injuries in the workplace in Great Britain.* Merseyside, UK: Health and Safety Executive.

Hubbard, D., & Evans, D. (2010). Problems with scoring methods and ordinal scales in risk assessment. *IBM Journal of Research and Development, 54*(3), 2:1-2:10. doi:10.1147/JRD.2010.2042914

IATA. (2015). *AIRLINE MAINTENANCE COST EXECUTIVE COMMENTARY. An Exclusive Benchmark Analysis (FY2014 data) by IATA's Maintenance Cost Task Force.* International Air Transport Association.

IATA. (2017). *Safety Report 2016.* Montreal: International Air Transport Association.

IBM. (2013). IBM SPSS Statistics for Windows version 22. Armonk, NY: IBM Corp.

ICAO. (1993). Human Factors Digest No. 7: Investigation of Human Factors in Accidents and Incidents. Circular 240-AN/144. Montréal: International Civil Aviation Organization.

ICAO. (2010). Annex 13 to the Convention on International Civil Aviation: Aircraft Accident and Incident Investigation. *10th*. Montréal, Canada: International Civil Aviation Organization.

ICAO. (2013a). Safety Management Manual. *Doc. 9859*. Montreal, Canada: International Civil Aviation Organization.

ICAO. (2013b). *State of Global Aviation Safety.* Montréal, Canada: International Civil Aviation Organization.

ICAO. (2015). Manual of Aircraft Accident and Incident Investigation. *Doc 9756 AN/965. Part I Organization and Planning, 2nd*. Montréal: International Civil Aviation Organization.

ICAO. (2017a). *ADREP Taxonomy*. Retrieved March 1, 2017, from http://www.icao.int/safety/airnavigation/AIG/Pages/ADREP-Taxonomies.aspx

ICAO. (2017b). *Safety Report.* Montreal: International Civil Aviation Organization.

IMO. (1997). RESOLUTION A.849(20) adopted on 27 November 1997 CODE FOR THE INVESTIGATION OF MARINE CASUALTIES AND INCIDENTS . International Maritime Organization.

IMO. (2015). Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process. MSC-MEPC.2/Circ.12/Rev.1. London: International Maritime Organization.

Jackson, C. A., & Earl, L. (2006). Prevalence of fatigue among commercial pilots. *Occupational Medicine, 56*(4), 263-268. doi:10.1093/occmed/kql021

Karanikas, N. (2015). An Introduction of Accidents' Classification Based on their Outcome Control. *Safety Science, 72*, 182-189. doi:10.1016/j.ssci.2014.09.006

Karanikas, N. (2017). Human Factors Science and Safety Engineering. Can the STAMP Model Serve in Establishing a Common Language? *Proceedings of the 32nd EAAP Conference, 26-30 September 2016* (pp. 132-149). Cascais, Portugal: European Association for Aviation Psychology.

Karanikas, N., & Kaspers, S. (2016). Do Experts Agree When Assessing Risks? An Empirical Study. *Proceedings of the 50th ESReDA Seminar 18-19 May 2016.* Seville: European Safety, Reliability and Data Association.

Kaspers, S., Karanikas, N., Roelen, A., Piric, S., van Aalst, R., & de Boer, R. J. (2016). Exploring the Diversity in Safety Measurement Practices: Empirical Results from Aviation. *Journal of Safety Studies, 2*(2), 18-29. doi:10.5296/jss.v2i2.10437

Kjellén, U. (2009). The safety measurement problem revisited. *Safety Science, 47*(4), 486-489. doi:10.1016/j.ssci.2008.07.023

Leveson, N. (2011). *Engineering a Safer World: Systems Thinking Applied to Safety.* Cambridge, MA, USA: MIT Press.

Li, G., Baker, S. P., Grabowski, J. G., & Rebok, G. W. (2001). Factors associated with pilot error in aviation crashes. *Aviation, Space, and Environmental Medicine, 72*(1), 52-58.

Li, W. C., & Harris, D. (2005). HFACS Analysis of ROC Air Force Aviation Accidents: reliability analysis and cross-cultural comparison. *International Journal of Applied 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, and Environmental medicine, 78*(4), 420-425.

Li, W. C., Harris, D., & Yu, C. S. (2008). Routes to failure: Analysis of 41 civil aviation accidents from the Republic of China using the human factors analysis and classification system. *Accident Analysis & Prevention, 40*(2), 426-434. doi:10.1016/j.aap.2007.07.011

Maurino, D. E., Reason, J., Johnston, N., & Lee, R. B. (2017). *Beyond aviation human factors: Safety in high technology systems.* New York: Routledge.

Nemeth, C., & Hollnagel, E. (2016). *Resilience Engineering in Practice, Volume 2: Becoming Resilient.* CRC Press.

O'Hare, D., Wiggins, M., Batt, R., & Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation. *Ergonomics, 37*(11), 1855-1869. doi:10.1080/00140139408964954

O'Hare, D. (2006). Cognitive functions and performance shaping factors in aviation accidents and incidents. *The International Journal of Aviation Psychology, 16*(2), 145-156. doi:10.1207/s15327108ijap1602_2

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381-410. doi:10.1177/0018720810376055

Patriarca, R., Bergström, J., Di Gravio, G., & Costantino, F. (2018). Resilience engineering: Current status of the research and future challenges. *Safety Science, 102*, 79-100. doi:10.1016/j.ssci.2017.10.005

Plioutsias, A., & Karanikas, N. (2015). Using STPA in the Evaluation of Fighter Pilots Training Programs. *3rd European STAMP Workshop. 128 (1)*, pp. 25-34. Amsterdam: Procedia Engineering. doi:10.1016/j.proeng.2015.11.501

Rasmussen, J. (1983). Skill, Rules, Knowledge: Signals, Signs and Symbols and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*(3), 257-266. doi:10.1109/TSMC.1983.6313160

Reason, J. (1990). *Human Error.* New York: Cambridge University Press.

Reason, J. (1998). Achieving a safe culture: theory and practice. *Work & Stress, 12*(3), 293-306.

Reason, J. (2008). *The human contribution: unsafe acts, accidents and heroic recoveries.* CRC Press.

Sarter, N. B., & Alexander, H. M. (2000). Error types and related error detection mechanisms in the aviation domain: An analysis of aviation safety reporting system incident reports. *The*

International Journal of Aviation Psychology, 10(2), 189-206. doi:10.1207/S15327108IJAP1002_5

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *The International Journal of Aviation Psychology, 2*(4), 303-321. doi:10.1207/s15327108ijap0204_5

Schvaneveldt, R. W., Beringer, D. B., & Lamonica, J. A. (2001). Priority and organization of information accessed by pilots in various phases of flight. *The International Journal of Aviation Psychology, 11*(3), 253-280. doi:10.1207/S15327108IJAP1103_02

Shappell, S. A., & Wiegmann, D. A. (2003a). Reshaping the way we look at general aviation accidents using the human factors analysis and classification system. Proceedings of the 12th International Symposium on Aviation Psychology. The Ohio State University.

Shappell, S. A., & Wiegmann, D. A. (2003b). *A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998. DOT/FAA/AM-03/4.* Washington: Civil Aeromedical Institution, Federal Aviation Administration.

Shappell, S. A., Detwiler, C. A., Holcomb, K. A., Hackworth, C. A., Boquet, A. J., & Wiegmann, D. A. (2006). *Human error and commercial aviation accidents: A comprehensive, fine-grained analysis using HFACS. DOT/FAA/AM-06/18.* Washington: Office of Aviation Medicine, Federal Aviation Administration.

Stamatelatos, M., & Dezfuli, H. (2011). *Probabilistic risk assessment procedures guide for NASA managers and practitioners. NASA/SP-2011-3421.* NASA, Office of Safety and Mission Assurance. Washington: National Aeronautics and Space Administration.

Telfer, R. A., & Moore, P. J. (1997). *Aviation training: Learners, instruction and organization.* New York: Routledge.

Tsang, P. S., & Vidulich, M. A. (2006). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (3rd ed., pp. 243-268). New Jersey: John Wiley & Sons. doi:10.1002/9781118131350.ch8

Verstraeten, J. G., Roelen, A. L., & Speijker, L. J. (2014). *Safety performance indicators for system of organizations in aviation.* Amsterdam: National Aerospace Laboratory NLR.

Wiegmann, D. A., & Shappell, S. A. (1997). Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. *The International Journal of Aviation Psychology, 7*(1), 67-81. doi:10.1207/s15327108ijap0701_4

Wiegmann, D. A., & Shappell, S. A. (2003). *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System.* Aldershot: Ashgate Publishing.

Wiggins, M. W., & Loveday, T. (2015). *Diagnostic Expertise in Organizational Enviroments.* Surrey, UK: Ashgate.

Wise, J. A., Hopkin, V. D., & Garland, D. J. (2009). *Handbook of aviation human factors* (2nd ed.). Florida: CRC Press, Taylor & Francis Group.

Woods, D. D., Dekker, S., Cook, R., Johannesen, L., & Sarter, N. (2010). *Behind Human Error* (2nd ed.). Farnham: Ashgate.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *State-of-the-Art Report Behind human error: Cognitive systems, computers and hindsight. No. CSERIAC-SOAR-94-01.* Alexandria, VA: Crew System Ergonomics Information Analysis Center, Defense Technical Information Center.

Appendix A: Generic examples of the controllability classification

CONDITIONS	SYSTEM	CONTROLLER	ACTION	CONTROLLABILITY	FAMILIARITY LEVEL	REACTION EFFECT
Aviation sector						
One tyre of the main landing gear system blew out during taxiing.	Aircraft during flight operation	Flight crew	The pilot flying stopped the aircraft immediately after the realisation of the failure.	Controlled	MHF	Positive
Engine flameout during landing approach under adverse weather conditions.	Aircraft during flight operation	Flight crew	The crew faced the failure and landed the aircraft safely.	Controlled	LMF (due to the adverse weather conditions)	Positive
Aircraft malfunction leads to an unstable state while in final approach and results in a runway excursion.	Aircraft during flight operation	Flight crew	Incorrect techniques resulted in a runway excursion	Controlled	LMF (instability of aircraft in a critical flight phase)	Negative
	Emergency services	Air Traffic Controller	The ATC mobilised the emergency services immediately	Controlled	MHF	Positive
	Emergency services	Fire brigade	The fire brigade personnel sprayed the engines and prevented fire.	Controlled	MHF	Positive
Severe impacts on the engine compressor blades due to Foreign Object Damages observed during After Flight Inspection.	Aircraft during flight operation	Flight crew	(Not applicable)	Uncontrolled (damages could have affected engine performance before landing, with an unforeseeable outcome)	(Not applicable)	(Not applicable)
	Aircraft under maintenance	Technician	The technician discovered the impacts during the inspection	Controlled	MHF	Positive
Other sectors/activities						
Engine fire while driving in the middle lane of a motorway with dense traffic.	Vehicle moving	Driver	The driver pulled the vehicle over safely and successfully extinguished the fire	Controlled	LMF (due to the original position of the car, the unexpected fire and the traffic around)	Positive
Engine oil light illumination during driving.	Vehicle moving	Driver	The driver saw the warning and continued driving the vehicle until the next car station; the engine suffered severe damages.	Controlled	MHF	Negative

CONDITIONS	SYSTEM	CONTROLLER	ACTION	CONTROLLABILITY	FAMILIARITY LEVEL	REACTION EFFECT
Ship manoeuvring in the port under adverse weather conditions and heavy traffic.	Ship moving	Ship crew	Ineffective manoeuvring resulted in a collision with another ship.	Controlled	LMF (due to the weather conditions and traffic)	Negative
Blood quality problems due to exposure to hazardous chemicals at the working place were identified after non-scheduled medical checks.	Body	Patient	(Not applicable)	Uncontrolled (the discovery of the problem was a matter of chance)	(Not applicable)	(Not applicable)
	Patient	Medical doctor	Suggested additional exams based on the health problems reported by the patient	Controlled	MHF	Positive
	Working environment	Employer of the patient	(Not applicable)	Uncontrolled (the employer had not identified risks related to exposure to chemicals)	(Not applicable)	(Not applicable)

Appendix B: Examples of application of the controllability classification to safety investigation reports studied

EVENT CLASS	CONDITIONS	SYSTEM	ACTION	CONTROLLABILITY	FAMILIARITY LEVEL	REACTION EFFECT
Accident	Multiple bird strikes led to mechanical damages in both engines during climb.	Aircraft during flight operation	The flight crew maintained control and managed to ditch and evacuate the aircraft without casualties.	Controlled	LMF	Positive
Accident	Aircraft crashed into high terrain about 3 miles southwest of the airport during a Non-Precision Approach.	Aircraft during flight operation	Due to combined effects of fatigue and inadequate training, the captain did not react properly to ground proximity warning callouts.	Controlled	LMF	Negative
Incident	Two aircraft were approaching the same destination in uncontrolled airspace.	Aircraft during flight operation	The flight crew of aircraft 1 were notified through its Airborne Collision Avoidance System (ACAS) of aircraft 2. Aircraft 1 performed a missed approach before a traffic alert, while aircraft 2 proceeded away from the approach area.	Controlled	MHF	Positive
Serious incident	In night conditions and under good visibility, the crew took off from a taxiway, unaware of the situation until the risk of the collision had passed.	Aircraft during ground operation	(No opportunity for intervening in the unfolding adverse conditions)	Uncontrolled	(Not applicable)	(Not applicable)

Appendix C: Variables and sample size

Group of event variables	Event variable (Sample size)	Values and their percentage in the sample					
Registration of the aircraft	Country (309)	A1 (21,0%)	A2 (11,7%)	A3 (10,0%)	A4 (26,9%)	A5 (17,8%)	Other (12,6%)
Temporal	Year (317)	1900-1999 (21,1%)	2000-2005 (28,7%)	2006-2009 (27,5%)	2010-2014 (22,7%)		
	Season (317)	Winter (28,4%)	Spring (20,8%)	Summer (29,7%)	Autumn (21,1%)		
	Daytime (302)	Morning (31,1%)	Afternoon (40,1%)	Evening (23,5%)	Night (5,3%)		
Aircraft characteristics	Age (301)	≤ 6 years (25,6%)	7-14 years (25,2%)	15-24 years (24,9%)	> 24 years (24,3%)		
	Type (312)	Jet (48,7%)	Propeller (37,2%)	Rotary (14,1%)			
	Weight (314)	≤ 27000Kg (57,3%)	> 27000Kg (42,7%)				
Operational characteristics	Category (306)	Commercial Air Transport (72,5%)	Other (27,5%)				
	Scope (302)	Passenger (54,3%)	Non-passenger (45,7%)				
	Phase (317)	En route (30,0%)	Ground (28,7%)	Other (41,3%)			
Event characteristics	Severity (317)	Accident (64,1%)	Serious incident (20,8%)	Incident (15,1%)			

Group of event variables	Event variable (Sample size)	Values and their percentage in the sample						
	Type (317)	Control Flight into Terrain (CFIT) (6,9%)	Loss of Control in Flight (LOC-I) (14,8%)	Mid-air Collision (MAC) (10,4%)	Runway Excursion (RE) (6,9%)	Power Failure (PF) (15,5%)	Other Technical Failure (OTF) (12,3%)	Other (33,1%)
Crew/ground staff fatigue	**Fatigue (317) [when stated as factor in the investigation report]**	Yes (9,8%)	No (90,2%)					