33RD CONGRESS
OF THE INTERNATIONAL COUNCIL
OF THE AERONAUTICAL SCIENCES
STOCKHOLM, SWEDEN, 4-9 SEPTEMBER, 2022

ICAS 2022 SWEDEN

STRESSING SAFETY ASSESSMENT METHODS BY HIGHER LEVELS OF AUTOMATION

Lothar Meyer[1], Christian Bjursten Carlsson[1], Åsa Svensson[2], Maximilian Peukert[1], Lars Danielson[3] & Billy Josefsson[1]

[1]LFV, Swedish Air Navigation Service, Research & Innovation, firstname.lastname@lfv.se
[2]LFV, Swedish Air Navigation Service, Research & Innovation, asa.e.svensson@lfv.se
[3]SDATS, Saab Digital Air Traffic Solutions, lars.danielson@saabgroup.com

Abstract

Automation aims to improve the system performance by reducing the workload for the operator, increasing the precision of the work tasks executed, enabling high reliability of the operations, and making sure the system is more efficient in performing the operations and in the end increasing safety. The side effect of higher levels of automation is increasing complexity of the socio-technical system that has the potential for automation surprise. In air traffic control, consequences of automation surprise may be safety-relevant. The question arises whether conventional safety assessment methods sufficiently support the safety assessor in evaluating the risk contribution of automation to prove the completeness of the assessment. This paper presents a preliminary review and problem analysis of conventional safety assessment methods used in safety-related work domains. The focus is on a systematic review of assessment in terms of both explicit and implicit support for detecting and mitigating the arousal of surprise effects. This is supplemented with a literature review of the factors that contribute to the occurrence of automation surprises. The working principles of safety assessment methods are contrasted with the characteristics of the occurrence of automation surprises. The results are discussed in the context of two conceptual shortcomings of assessment methods: The increased complexity and resulting error propagation, and the effects of linear model assumptions. Finally, solutions are briefly proposed that could support future development of valuation methods. .

Keywords: Human Factors, Safety Assessment, Automation Surprise, Problem Analysis, Socio-Technical Systems

1. Introduction

Higher level automation is a key enabler in Air Traffic Control (ATC) to help achieve the goal of increasing productivity and cost efficiency while improving safety [1]. Numerous automation tools are already implemented in operations to support the Air Traffic Control Operator (ATCO). En-route ATCOs use a number of automated tools to help identify and resolve conflicts, thus forming parts of the safety net. One example is the technical instrument Advanced Surface Movement Guidance and Control System (A-SMGCS) in tower control. It provides automatic route guidance to the ATCO based on parameters such as runway configuration and other constraints. The aircraft is then guided along an approved route by switching taxiway lights and stop barriers on and off. At the highest level of implementation, A-SMGCS can also issue conflict warnings when an aircraft and other movements on or near the runway come into conflict. Today, A-SMGCS is implemented as Surveillance Service and Airport Safety Support Service [2].

Another example of advancing digitalization is the concept of remotely controlled and digital tower (RTC), a redesign of the existing conventional tower workplace. RTC provides a digital platform that is capable of implementing more automation than the conventional tower, such as radar labels in the visual presentation system. When operating one airport, RTC is considered a supporting tool with its

digitalization. For example, new automation tools such as "visual tracking" allow the ATCO to quickly identify the aircraft using a visual overlay indicator on the visual presentation, replacing the out-the-window view. More advanced operational concept of controlling multiple airports (Multi Remote Tower), i.e. three airports simultaneously with only one ATCO, implies a significant transformation of working procedures and methods and also a need for more automation.

Viewed as a whole, the current level of automation is generally low in ATC, rather providing additional information to the ATCO than suggesting solutions or executing them. The reason for the rather automation-resistant approach in ATC is the acceptance by operators and the safety performance record of human operators so far. In 2020, no Air Traffic Management or Air Navigation Service-contributed (ATM/ANS, including ATC) fatal accidents or serious incidents within the EASA Member States [3], which corresponds statistically to 0.0 fatal accident and 0.7 serious incidents per million instrument-ruled (IFR) flights. In the last 10 years, there have been no fatal accidents with ATM/ANS contribution, which represents an all-time low in ATM history.

Nevertheless, workload and cost-efficiency arguments, stated by the European Union's Single European Sky ATM Research (SESAR) project, are of increasing interest and push the trend toward automated decision-support [3]. Implementing automation while maintaining the same (equal) level of safety (ELOS) remains the ultimate goal.

The reason to consider this a challenge is the lack of practical experience and historical data, making it difficult to quantify the impact of a new automated system on safety. The problem becomes more serious the larger the innovation step and the larger the blind spots [4]. An inevitable side-effect of these blind spots is the triggering of surprise effects in human collaboration with automation, "..when the automation behaves in a manner that is different from what the operator is expecting.", which is called Automation Surprise (AS) [5] or Automation Startle [6]. Automation Startle can be considered an automatic, physiological response triggered by a sudden event that contradicts the operator's expectations. However, AS can be seen as a cognitive-emotional response to something unexpected, resulting from a discrepancy between expectations and perceptions of the environment [7].

AS has proven to be a relevant safety issue to aviation [8, 9], and has the potential to diminish human safety performance during time and safety-critical situations. Although automation is supposed to increase safety and reduce workload, it might only take one surprise to turn the tide and wipe out the safety record achieved so far, as described by the automation paradoxon [10, 11]. Today, AS is mainly experienced in highly automated workplaces such as the flight deck, which plays a pioneering role in the use of highly automated aids. Concerning AS, the US Federal Aviation Administration (FAA) published a report of the Flight Deck Automation Working Group [12]: "The occurrence of Flight Management System (FMS) programming errors and lack of understanding about FMS operation appear across age groups and cultures. The errors are noteworthy and it has not been possible to mitigate them completely through training (although training could be improved). This reflects that these are complex systems and that other mitigations are necessary." One reason for the persistent problem of AS are Safety-I working principles, which are proving inadequate "to predict all failures in advance, find, and eliminate all the causes for them" as systems become more complex [13]. Safety-II working principles might be better suited to mitigate AS, but still lack the ability to prove a system for ELOS. The key to connect AS into Safety-I working principles seems to be a better understanding of the causes of AS and adjusting, or improving, the models of accident causation used in safety assessment methods accordingly. AS is seen as the biggest bottleneck that needs to be overcome to pave the way to higher levels of automation [10, 13].

In light of this challenge, the purpose of this paper is to investigate the existing methodological implications for providing a prospective safety assessment about whether automation and related operational concepts are at least equally safe as prior to implementation. The problem analysis presented here shall provide a better understanding of the gaps and limitations in conventional safety assessment methods. In this way, the analysis contributes to future methodological developments in the field of prospective safety assessment, to avoid the occurrence of AS or to reduce its probability.

The tracing of the discrepancy between the conclusions made about AS and the methodological assumptions of accident causation for safety assessments will shape the choice of structure and approach of this preliminary literature review. The state of research on AS assessment problems

is first presented, including an overview of the AS phenomenon. The research questions, aimed at finding the gap between AS-causing factors and the state of safety assessment, are presented below. In the methods section, our criteria for identifying literature suitable for providing insights into the questions outlined in the previous section are presented. The results are then presented and discussed. Finally, a conclusion shall be drawn, providing explanations on the problem of identifying, assessing and mitigating AS-related safety events.

2. Related Work

Introducing higher levels of automation in high-reliability organizations (HRO) where humans still have the legal responsibility can cause AS [14, 15]. AS is a mismatch between the automation's actions (or none actions) and the human's awareness and interpretation of the automation. AS and the lack of handling it can result in various forms of risks. There are many factors thought to be associated with AS, such as the concept of mode error, where the operator acts on the assumption that a particular mode is present, when in fact the automation is operating in a different mode [16]. Other factors are technical failures, operator's lack of mode awareness, misunderstandings / misinterpretations [17] by the operator of the automation, automation bias, automation complacency, and distrust / over trust.
The use of the concept of situational awareness to explain AS is controversial [18]. Nevertheless, the ability to interpret and predict unfolding events [19] is very critical and important. However, the operator's situation awareness can be negatively impacted due to AS [20]. In addition, lack of situation awareness can also lead to AS [21]. Hence, situation awareness and ASare related processes affecting each other. Even though automation is implemented to support the human operator, often with the purpose to increase efficiency and safety, automation can also contribute to a loss of safety if not implemented in the right way. However, finding the blind spots [22] of the automation, knowing when or where it might fail the human operator, is difficult to identify.
Many attempts have been made through Safety-I thinking [23] (basic safety assessment methods). However, such methods focus more on static systems and already happened incidents, leaving little room for improvement before it is too late. Prospective safety assessment methods aim at identifying and mitigating causes of harmful events during the pre-implementation phase. As this step is before startup to operations, causal models are used to interconnect causal relations between system functions, software- and hardware design, and operators. Reviews, focusing on similar objectives, pointed out the missing human factors component. Levson points out 2004 [24] that system safety used to be based on experience-based maturation of system design, whereas today's system design lacks experience due to constant change. A rising level of complexity between human and automation leads to new types of errors that result e.g. from inadequacies in the communication between them. Brooker 2010 [25] addresses implications of safety decision-making concerning the implementation of SESAR-related innovation. He proposes using human-in-the-loop simulations to prove ATM's ability to withstand disruptive events and prove an equal level of safety of the foreseen innovation. As long as the Safety-I methods are used under the recognition of their limitations, Safety-I methods can be effective. Safety-I methods can provide useful information and improve system and training design and processes, and safety culture. Safety-I methods are prospective, meaning that they are based on accident models that describe expected hazards and their effects on operations. However, the predictive power is subject to model inaccuracies and leaves a residual risk of an accident occurring. It is, thus, important to acknowledge and understand the processes and situation of the accident to mitigate the negative effects and prevent its recurrence. Because of increasing interest in alternative approaches, Hollnagel and Woods suggest focusing on joint cognitive systems [26] (human-machine collaboration in complex environments) since unpredicted and dynamic events unfold rapidly with automation (especially in the ATC). In recent years, the Safety-II concept [23] and Resilience Engineering [27] have come to complement safety assessment methods. They are intended to help reduce surprises in automation by examining what is running correctly and understanding normal variability in performance and execution. This approach complements Safety-I by eliminating and preventing hazardous situations, rather than identifying the root causes of what went wrong. However, automation surprises still happen [28, 21, 14] and it is of importance to further merge the Safety-II thinking and resilience with safety assessment methods for dynamic human-automation operations.

3. Research Questions

In the light of AS and the research literature mentioned above, this paper will explore the following two research questions:

1. Which factors are evoked by higher levels of automation that contribute to the occurrence of AS?

2. To what extent do current predictive safety assessment methods take into account AS and related concepts?

The first question aims to understand the way in which pattern and combination of situation factors emerge at higher levels of automation that are interrelated to the occurrence of AS. The second question has the purpose to review the safety assessment methods on the capability to cope with the challenges of AS. This concerns primarily the support that is provided by the method that helps the safety analysts or applicants of the method to identify and assess factors on their contributing to AS and mitigate them if needed.

4. Method

The chosen approach is to investigate literature that supports the understanding and problem analysis of the AS phenomenon and by this gives answers addressing our research questions above. The literature review focuses on the theories and concepts available in the respective research area of AS. The results of the review shall be synthesized qualitatively according to our questions. This approach is hence consistent with a meta-synthesis. The overview does not claim to be exhaustive, but is preliminary and representative based on a selection from acknowledged databases and the corresponding number of citations, as well as on the judgment of the authors that have experience in the field of safety assessment, accident investigation, and human factors.
The first research question is intended to be answered based on literature reviews that satisfy the criterion of developing concepts for explaining AS. This may involve higher-level causal classes, terms, taxonomies, and models. Findings that suit providing an answer are finally derived from experienced cases of AS (inductive). For simplicity, the search focuses on cockpit crew experience and cockpit automation, as this research area is at the leading edge and provides the most experience. Qualitative studies of interest may also aim to develop concepts and discussions, which in turn may be based on literature reviews and logical reasoning. Regarding the source of the cases, experience-based data from operators involved, such as surveys or interviews, are subjective but externally valid and provide details about the working context of the particular situation in which AS occurred. Additionally, the literature review may also include investigation reports from safety occurrences, involving accidents and incidents. These cases rely hence on e.g. flight data recorder, cockpit voice recorder as well as pilots testimonies. Factors contributing to AS shall be identified that represent an intersection of the findings from the literature reviewed. In order to structure the factors and the discussion based upon it, a classification system shall be applied. The chosen classification system SOAM from Eurocontrol [29] is a well-known system usually used in accident and incident investigations and is well suited for our approach.
The second question will be approached using a preliminary review of safety assessment methods. We define a safety assessment method as a framework of methods, tools, procedures, models, and techniques to develop safety assessments that support the change of functional systems by identifying riskful events, mitigating risks where needed and determining whether the target level of safety is met. The characteristics of methods shall be investigated, giving explanation on how AS is given space to unfold even though the method was applied appropriately. The review shall also cover related concepts, such as situation awareness and mode awareness that are closely related to AS. As assumed terminology evolved over the decades, AS-related terms might be addressed implicitly by their corresponding predecessors. In the scope of this literature review, we search for those terms and concepts that cover the contribution of the human-automation interaction to accident causation. This shall compensate for the development and refinement over the decades that concepts have undergone, such as “human error” and “unsafe act”.

5. Results

5.1 Reviewing Factors Contributing to Automation Surprise

The search of the term “automation surprise” in title, keywords, and abstract resulted in 86 hits on the Scopus search platform. The additional narrowing with the keyword “safety” in all-text led to 68 hits. Further filtering was applied according to the targeted profile lined out in the method section further reduced the set of hits to 5. Additionally 6 studies were found through citation that matched the target criteria and was therefore added to the selected list.

Sherry & Mauro 2014 [30] investigated 19 loss-of-control aviation accidents with a focus to understand the sequence of events that coincided with inadequate intervention actions by the flight deck crew. They found that the decision-making logic was not adequately supported by the automation cues available in the cockpit, which are necessary for an appropriate response. The cues of concern are aircraft structure and airfoils, aircraft sensors, control surface, propulsion systems, and the automation itself. Rare failure events that are related to the cues were not detected in the accidents due to the “hidden” nature of fail-safe sensor logic, “silent” and/or “masked” automation responses. Additionally, it was mentioned that cues are absent that support the pilots to anticipate speed envelope violation, in recognizing a speed envelope violation due to noise in the airspeed signal, in recognizing the airspeed envelope violation due to non-linearity and latency in the thrust response near the idle thrust setting.

Dehais & Peysakhovich 2015 [31] provoked the analyzed response of the flight deck crew to automation surprise in a flight simulation using eye-tracking. Indeed, “automation surprise” led to an excessive but inefficient visual search that prevented pilots from extracting the relevant information (i.e. the speed indicator). Whereas conflict solving was “straightforward” (i.e. reducing the selected speed with the dedicated FCU knob), most pilots were stuck and failed to deal with the situation immediately. Many participants made typical “fixation errors” as they persisted in disengaging and reengaging the autopilot (i.e. lateral/vertical guidance) or dialing in vain the altitude or vertical speed knobs on the FCU. Moreover, the analysis of ocular events revealed that the volunteers exhibited higher visual search (more short fixations and saccades) to the detriment of information processing (fewer fixations) during conflict in comparison to baseline. The analysis of eye movement revealed that such conflicts impair attentional abilities, leading to an excessive visual search and inability to extract relevant information.

Rankin, Woltjer & Field identified, with the help of 20 interviewed pilots in 2019 [32], 9 categories on the basis of 48 cases of automation surprise experienced in cockpits. The causes investigated were an absence of salient cues, causing confusion of switches. Pilots found it difficult to detect passive and insidious disturbances that build up slowly over time that make the autopilot suddenly disconnect. As well, it was found time-critical to deal with conflicting and inconsistent data from multiple failures.

Boer and Hurts investigated 2017 [33] using a survey of flight deck crews that the occurrence of AS occurs 3 times per pilot and year without any severe consequences. They found AS events not to be the result of cognitive failures, but rather the consequence of the current complexity of the cockpit system and interface design choices that possibly exceed the bounds of human comprehension. The effect of experience and operational intensity indicates that the initial training curriculum for pilots is insufficient to avoid AS events. Given that system complexity and interface design choices are a major factor in so many (non-consequential) AS events, it seems that this overrides individual cognitive errors and differences in knowledge and training.

In 2017, Boer and Decker [14] examined the theories of AS obtained to date. They compared two models that explain the occurrence of AS, identifying a common pattern of occurrence. First, automated systems act on their own without immediately preceding instructions, input, or commands from human(s). Second, there are gaps in users’ mental models of how automation works. And third, feedback about automation activities and future behavior is weak.

Parasuraman and Mazey reviewed in 2010 [34] the different phenomenons of automation complacency and automation bias, prerequisites to the arousal of AS. Automation complacency arises primarily in the attention allocation strategy of keeping track of parallel tasks in a mixed environment of automated and manual work. Attention is preferably shifting to manual work at the expense of monitoring automation. Automation bias aims at omission and commission error when decision aid

is wrong, evoked by the operator's attitude or assumption of relying on flawless automation.
Endsley notes in 2016 [35] that situational awareness can be negatively affected by automation. A major factor in automation-related errors is the operator's lack of awareness of the state of automation. The lack of awareness is indicated by a slow detection of problems that comes along with extra time needed to understand relevant system parameters and settings. The cause might be a loss of vigilance and increasing complacency, rather taking up a position of monitoring and passively receiving information than processing and anticipating in the scope of appropriate situation awareness.
Sarter & Woods explained 1997 [36] that automation surprises are indications that a crew has misunderstood, miscommunicated with, misused, or mismanaged the automated systems. They distinguish two different types of AS:

- automation does not execute actions that were expected and
- automation does change inputs or executes in a different manner than it was told by the operator.

Sarter, Woods and Billings 1997 [15] argued that the gap between user-centered intentions and technology-centered development raises the likelihood for AS arousal. AS likely occurs when

- designers oversimplify the pressures and task demands from the users' perspective,
- assuming that people can and will call to mind all relevant knowledge,
- overconfidence that they have taken into account all meaningful circumstances and scenarios,
- assuming that machines never err,
- making assumptions about how technology impacts human performance without checking for empirical support or despite contrary evidence,
- defining design decisions in terms of what it takes to get the technology to work,
- sacrificing user-oriented aspects first when tradeoffs arise and
- focus on building the system first, then trying to integrate the results with users.

Further, AS is not simply the result of over-automation or human error. Instead, they represent a failure to design a coordinated team effort across human and machine agents as one cooperative system.
Sarter & Woods summarizes 2000 [37] that the potential for automation surprise is greatest in the following cases: (1) automated systems act on their own without immediate preceding directions from their human partner, (2) gaps in user's mental models of how their machine partners work in different situations, and (3) week feedback about the activities and future behavior of the agent relative to the state of the world.
Decker provides a list of circumstances in 2002 [11] under which AS is likely to occur. The automation may be undergoing a mode change from someone who programmed it, or a while ago, or the automation may be following a pre-programmed logic. There is insufficient feedback about its behavior; instead, the automation tends to communicate a status to the user. Event-driven circumstances can create situations where the automation dictates to the user how quickly to think, decide, and act. It may be difficult for the user to assess what input is required for the automation to do what the user wants.

5.2 List of Factors Contributing to Automation Surprise

From the preliminary review, 13 factors could be identified and classified using Eurocontrol SOAM (see Table 1).

Table 1 – AS-contributing factors and related to SOAM elements

Factors	SOAM Catergory	SOAM Element
Out-the-loop effect (low vigilance, incomplete or corrupted situation awareness)	Human Performance Limitation	Contextual Conditions
Lack (“masked”/”silent”-mode change) of or excessive feedback from system	Workplace Conditions	
Contradictory feedback from system	Workplace Conditions	
Automation complacency	Human Performance Limitation	
Automation bias / Overtrust	Attitudes and personality factors	
Fatigue	Physiological and emotional factors	
Low workload and High workload	Workplace Conditions	
Complexity (unmanageable number of dependencies between the operator- automation and automation-automation)	Human Performance Limitation	
Poor understanding of automation working principles	Human Performance Limitation	
Poor training in the handling of automation	Human Performance Limitation	
Gap between technology-centered design and human-centered design	Equipment and Infrastructure	Organization
Technical-related breakdown/degradation of automation level	Maintenance Management	
Poor automation design	Equipment and Infrastructure	

5.3 Review Safety Assessment Methods

The preliminary literature search yielded 16 methods briefly presented in chronologically ascending order below. The methods provide a representative cross-section of the most prevalent safety assessment methods and thus have no claim to be exhaustive. The chronological distribution of the presented methods shall reflect the method development process across the years, starting in 1949 with the Failure Mode and Effect Analysis method. Earlier methods such as Heinrichs' accident triangle of 1931 **??** were not considered in this preliminary review because we assume that the findings were transferred to later generations of methods and are thus implicit there.

The Failure Mode and Effects Analysis (FMEA) was established in 1949 [38, 39]. In the FMEA, the analysis is carried out inductively (from the bottom up). This is, in principle, in the reverse order in comparison with the fault tree analysis (FTA) model described below. The FMEA is based on components or subsystems for which each fault type is analyzed concerning the effect it can have on the system. The model requires specialist knowledge of the reviewed system. It is a structured method to find weaknesses in the system. The technique is detailed and entails a risk of missing overall disturbances. The method focuses on components and does not explicitly address automation surprise (AS) or situation awareness (SA).

Closely related to the FMEA is the Failure Mode, Effects, and Criticality Analysis (FMECA) model from 1949 [38, 39]. Analogous to qualitative fault trees, a risk matrix can be used to assess the various error types. The Failure Mode, Effects and Criticality Analysis introduces a column for severity, also called criticality. To fully utilize the model requires specialist knowledge of the system being reviewed. The method focuses on components and does not explicitly address automation surprise or situation awareness.

Human Reliability Analysis (HRA) was initiated in 1952 and originated in the nuclear industry [38, 39, 40]. The HRA is a generic name for methods to assess factors that may impact human reliability in probabilistic risk analysis to operate a sociotechnical system. There are many different methods with varying complexity. They are all based on the same underlying principle (Swiss cheese model). The HRA is a structured approach to identifying potential human failure events (HFEs) and systematically estimating the error probability using data, models, or expert judgment. The method provides no clear view of SA but indirectly by using performance shaping factors (PSF) for human activities. The method is based on expert knowledge, and the method does not explicitly address AS.

The HRA is associated with Probabilistic Risk Assessment (PRA) [38, 39]. This model is based on FTA/ETA and was established in 1965. The model is almost exclusively used in the nuclear industry. As with HRA, the model indirectly addresses since automation surprise can be addressed as an unwanted event as the starting point. The method does not explicitly address automation surprise or situation awareness.

The Fault Trees Analysis (FTA) model was developed in 1962 [38, 39]. The model identifies collab-

orative events that can lead to harmful events. The method is mainly used if the end consequence is severe. However, the harmful events are challenging to model correctly. The model is easy to overlook combinations and can thus give an incorrect image. Fault trees can be complex and challenging to see. Difficult to find relevant data if the error tree encompasses quantification. Properly executed, possible combinations and weak links are found in the system. The model can also give an idea of how common an event can be. The model addresses indirect automation surprise since that event can be approached as an unwanted event as the starting point. However, AS is not explicitly mentioned in the method; the same is valid for SA.

Preliminary Hazard Analysis (PHA) is from 1969 [38, 39]. The model identifies harmful events with a focus on high-level events. Reported as risk levels, the model is often performed as an initial analysis to screen preliminary hazards. The model provides limited identification of causes. AS is addressed Indirectly since AS can be addressed as an unwanted event as the starting point. However, AS is not explicitly mentioned in the method; the same is valid for SA.

The Hazard and Operability (HAZOP) study was established in 1974 [38, 39]. HAZOP is a systematic, team-based approach to assessing process hazards and potential operability issues. HAZOP identifies hazards and potential operability problems in a system or process commensurate with the level of detail available and generates a set of actions to help eliminate or minimize these. The model requires specialist knowledge of the systems being analyzed. This method is best suited for processes and operational processes. It is closely related to the FMEA method, but the error modes have been defined and made uniform from the beginning. AS can be addressed based on specialist knowledge but not directly supported by the method. The same is valid for SA.

Hazard Identification (HAZID) was introduced in 1993 [38, 39]. The model is a modification of the HAZOP model, especially to be used to identify human failures. HAZID is a systematic, team-based approach to identifying hazards and their potential consequences. HAZID is used at different stages in a project or lifecycle of a system, including the operations phase. It is commonly used to identify safety, health, and environmental hazards early in a project to help develop inherently safer design alternatives and to help guide future risk reduction activities. HAZID identifies hazards in a system or process commensurate with the level of detail available and generates a set of recommendations and actions to help eliminate or minimize identified hazards. The outcome from a HAZID study should be documented as a HAZID report, and actions followed up and closed out. AS can be addressed based on specialist knowledge and not directly from the method. SA is not explicitly managed.

The Event Tree Analysis (ETA) model was introduced in 1974 [38, 39], and the model identifies how a top event can escalate to possible final events. Event trees can be complex and challenging to get an overview of. Indicates possible and probable end events. With simple means, a rough quantification can be made to support the assessment of probability in risk assessment. The model addresses AS indirectly since AS can be approached as the starting point of an unwanted event. However, not mentioned explicitly in the method, which is also valid for SA.

The Bowtie model was introduced in 1979 [38, 39]. The Bowtie model is based on fault trees and event trees combined. The Bowtie method analyzes a hazard or a critical event through cause-consequence analysis. The left-hand side of the Bowtie is formed by a Fault Tree, which models how combinations of primary events cause the hazard. The right-hand side of the Bowtie is based on an Event Tree, which models the consequences of the hazard. Fault Tree Analysis (FTA) and Event Trees (ETA) are based on linear cause and effect trajectories. The model can address AS indirectly by utilizing AS as the starting point for an unwanted event; however, AS is not explicitly mentioned in the method. Likewise, SA is not used in the model.

In 1997 the model Barrier Analysis or Layers of protection analysis (LOPA) was introduced [38, 39, 41]. LOPA is a scenario-based risk analysis and can be said to be a simplified form of error trees and event trees. Hence the model is connected to the Bowtie model. LOPA is based on quantifying frequencies, risk reduction factors, and probabilities for relevant deviations, barriers, prerequisites, and escalation factors. The method is used to assess and report whether the barrier measures taken are sufficient—a structured study of the existing barriers and the opportunity to assess the reliability of the current application. The model poses difficulties in finding relevant data on both initiating event frequencies and component error probabilities. This can lead to both over-and under-evaluation of

scenarios. Either AS or SA is utilized in the model.
Safety Assessment Method (SAM) from 2000 [42] is based on the following phases: Functional Hazard Analysis (FHA), which identifies hazards and assesses their effects and the related severity; Preliminary System Safety Assessment (PSSA) which includes fault tree analysis, event tree analysis, common cause analysis; System Safety Assessment (SSA) which provides for documentation of the evidence, collecting data, test, and validation. SAM is a framework that contains methods and techniques to develop safety assessment of changes to functional systems for Air Navigation Service Providers (ANSP). SAM presents a general overview of Air Navigation Systems safety assessment from an engineering perspective. The model focuses on engineering issues and not primarily on changes in the functional system. The model is based on both Fault Trees and Event Trees. AS can be addressed indirectly as a starting point of an unwanted event. The same is valid for SA.
Cause Consequence Analysis (CCA) encloses the application of the cause–consequence diagram method to static systems [38, 39]. The model was introduced in 2002 and aimed to model, in diagrammatic form, the sequence of events that can develop in a system due to combinations of basic events. Cause Consequence Analysis combines bottom-up and top-down analysis techniques of binary decision diagrams (BDD) and fault trees. The result is the development of potential accident scenarios (and hence not necessarily used in Safety Assessments). Neither AS nor SA is mentioned in the model.
Systems Theoretic Process Analysis (STPA) was outlined in 2012 [38, 39, 43]. STPA is a qualitative hazard analysis technique that assumes that accidents occur not simply because of component failures but because constraints on component behavior are inadequately enforced. It is used to identify instances of inadequate control that could lead to the presence of hazards, to identify safety-related constraints necessary to ensure acceptable risk, and to gain insight into how those constraints may be violated. This information can control, eliminate, and mitigate system design and operation hazards. STPA can be applied to existing designs or proactively to help guide the design and system development. The model is based on the Systems-Theoretic Accident Model and Processes (STAMP). Overview of the model includes control actions provided to affect a controlled process, feedback may be used to monitor the process, process model (beliefs) formed based on feedback and other information, and the control algorithm determines appropriate control actions given current thoughts. AS is mentioned primarily since the model is based on control theory. SA is not explicitly mentioned in the model but can be identified as a hazard to the system.
Specific operations risk assessment (SORA) from 2019 [39, 44] is a multi-stage process of risk assessment aiming at risk analysis of certain unmanned aircraft operations and defining necessary mitigations and robustness levels. The model is aimed explicitly at unmanned aircraft and is based on traditional risk assessment methodologies—a qualitative method where AS can indirectly be addressed. SA is not addressed in the model.
High-fidelity risk modeling (HFRM) from 2022 [44]. While SOAR is a qualitative high-fidelity risk modeling, HFRM is a quantitative estimate regarding the operation's expected fatality rate (EFR). Neither AS nor SA was directly addressed. Neither AS nor SA is directly addressed. However, both AS and SA can be addressed as unexpected events to model the system behavior.

5.4 Supplementary Frameworks

Regulation EU 2017/373 (2020) [45] is not a safety assessment method. The regulation aims to identify changes in the functional system from a safety perspective. Hence, in the safety assessment of functional systems, it may not always be possible or desirable to specify safety criteria in terms of quantitative risk values. Instead, safety criteria may be defined in terms of other measures related to risk. These measures are called proxies that indirectly are the measure of risk. The 373 provides the opportunity to instead use risk analysis (for example, traditional safety methods like SAM) in terms of safety risks or the use of safety risks in terms of proxies. The regulation separates ATS providers (who need to complete Safety Assessments) and non-ATS providers (who need to develop Support Safety Assessments). Neither AS nor SA is explicitly mentioned in the model. It can, however, be addressed as failure modes that can initiate an unwanted event.
The Functional Resonance Analysis Method (FRAM) [46] was introduced in 2012. FRAM is a system-based method developed to understand complex sociotechnical systems. FRAM focuses on learning

from safety occurrences or undesirable states and can be utilized to understand how things go well in a system. This is encompassed by identifying gaps between work as imagined (WAI) and work as done (WAD). FRAM is used to model the functions needed for everyday performance to succeed and can then be used to explain specific events by showing how functions can be coupled and how the variability of everyday performance sometimes may lead to unexpected and out-of-scale outcomes - either good or bad. The FRAM is based on the four basic principles: equivalence of successes and failures, approximate adjustments, emergence, and functional resonance.

5.5 Summary

Section 5.3 shows 17 assessment methods, with 15 not referring to any explicit help to identify automation surprise or situational awareness issues. These 15 rely on high-level model assumptions that require the safety analysts to define events independently. Using event criteria that involve human error and implications of situational awareness is the analyst's choice. In addition, methods give no procedural guidelines that add support to conducting empirical studies such as historical data or human-in-the-loop simulations. In the majority, the judgment on the probability of events and effects on operations is based on expert knowledge. Leveson's STPA method is an exception, which relies on a control model framework that defines operator and automation in a continuous loop. The analyst uses the framework to model interaction elements that pass the interface between the operator and automation. Concerning supplementary frameworks, FRAM can also model events associated with automation surprise and situational awareness. The Human Reliability Analysis represents the franchising of methods that rely on predefined error classes applied to human perception, decisions, and actions.

6. Discussion

The present study has contributed with a preliminary literature review on the factors contributing to the occurrence of AS and the related concepts of explanation. A picture of the state-of-the-art knowledge can thus be compiled, providing the prerequisites to conclude on causative pattern and the change induced by raising levels of automation. Additionally, a literature review of safety assessment methods reveal the problems methods face to deal with AS.

6.1 Contributing Factors of Automation Surprise

6.1.1 List of Factors

There were 11 studies reviewed that fulfilled the criteria defined in the method section. The studies were conducted in the field of aviation and addressed concepts related to the common subject of AS. The review led to 13 factors presented in Table 1. The first finding after the review is the wide range of context in which factors were interpreted by the authors. Some studies locate the search in the context of operating procedures and input-output processes between operator and automation. On the other extreme, conclusions were drawn on a high- level that addressed cultural and organizational factors. The specific focus might vary strongly depending on the context of the respective study and authors. The second finding is that factors seem to be quite wide-scattered across the barriers and categories used by SOAM, not allowing for identifying the one and only root cause of interest. The results of the literature review are presented in Table 1, using aggregate terms intended to cover the terms defined or identified in the literature reviewed.

6.1.2 A SOAM approach to classification

Based on the investigation results, it remains unclear to what extent the factors are interrelated and interdependent to cause AS. The history of accident investigation suggests that factors rarely occur alone but it is the combinations that cause extreme circumstances that lead to AS. Depending on their order, in the escalation chain, the factors may be more directly or indirectly related to the occurrence of AS. The variety of order exhibits similarity to accident barrier models where organization and culture are understood as a barrier, just as the actions undertaken by the operator, committing the unsafe act. In both cases, it is an undesired event that is to be explained and avoided, on the one hand, the AS event, and on the other hand the accident event in the Safety I world.

We chose the Safety Occurrence Assessment Method [29] to classify the factors according to an established accident investigation scheme. This is to see the characteristics of their distribution along the escalation chain and to identify significant densification points, as shown in Table 1. Because AS induces a state of confusion that limits human performance, it may be best suited for the contextual barrier occurring immediately before a human error occurs.
The factors could be interrelated, as multiple barriers may be breached when AS occurs. For example, automation might give contradictory feedback because of poor automation design. This involves the categories “workplace conditions” as well as the “equipment and infrastructur” in different elements.
According to the SOAM classification result, the majority of factors appear to relate to “contextual conditions” (10 of 13), with a focus on “human performance limitations” (5 of 10) and “workplace characteristics” (3 of 10). This proves that AS is a phenomenon that emphasizes the limits of human performance at the edges of what he/she is capable of handling. The distribution within the contextual condition also shows that AS seems to be a phenomenon with scattered features, showing diverse signs of occurrence. At the organizational level, a key issue seems to be how use cases are designed and implemented by automation designers so that the automation works according to desired rules and pre-programmed logic.
The diverse character of AS inhibits the effective handling by safety management systems as observed pattern do not exhibit the high discriminatory power and significance to develop appropriate mitigations and reevaluate the automation accordingly. The characteristics and causes of AS are still too unclear today.

6.1.3 Limitations of AS investigation

Since AS is a phenomenon whose existence is based on subjective observations and reports of experience, there is little evidence to verify (e.g., by empirical means) whether the factors found are causes, side effects, or even consequences of AS. Accident investigations feature particularly extensive analysis of the flight data recorder and voice recorder, such as Flight AF447 in 2009 or TK1951 in 2009, based on which detailed conclusions can be drawn that refine understanding about AS. See e.g. Dekker (2009) [47]. Basically, natural observations are not able to draw cause-and-effect conclusions. This concerns particular conclusions based on the operator’s memory, the recalled scan pattern of the environment and behavior before, during and after the AS occurrence. Rather, it is the context of the experts’ operational experience and expectations that gives the factors a presumed "cause" status, which in research is also called a cause hypothesis, which is up for verification. The dilemma, however, in using exactly this method is the small size of the available samples, which face a comparatively high variability in the appearance characteristics of AS. The reliability of the data supplied, on which the conclusions are based, remains quite low. In this light, AS has to be seen as a term that covers a quite wide range of event characteristics that the operator may experience as surprising. It is likely to have many factors outside those that could have been determined, observed, and summarized in Table 1, located in the tail of the broad distribution curve. It is therefore likely that there are more gaps in the barriers that allow AS without any of the factors listed in Table 1 being of relevance.
Another perspective of AS investigations is the accidents of Lion Air flight 610 (JT610) 2018 [48] and Ethiopian Airlines flight 302 (ET302) 2019 [49] with the Boeing 737 Max 8 airplanes. In the wake of the accidents, the investigations revealed a novel type of automation surprise to the airline industry. In the Boeing 737 Max, the manufacturer had installed a new software system (Maneuvering Characteristics Augmentation System (MCAS)) to remedy hardware problems with the new 737 Max-series to make sure the new model looked and behaved like an ordinary Boeing 737 Next Generation (NG). On top of this, the MCAS received little or no attention for pilots during familiarization training when transferring from the NG-series to the Max-series. The pilots were unaware of the MCAS system and its effects when activated. In addition, there was no direct alarm indication to the pilots of any prevailing system failure when MCAS was activated. The closest non-normal situation was related to the uncommanded stabilizer trim movement, which led the pilots to assess the situation as a “Runaway Stabilizer” anomaly with connected actions. However, the instigation of MCAS led to a

situation beyond automation surprise since the pilots were surprised by the automation of a system they did not know was installed, which aggravated the outcome.
The verification of AS characteristic pattern by means of human-in-the-loop simulation is an appropriate response that aims at mitigating the subjectivity of operator testimonies. The flip side of the coin is that one must rely on a relatively high maturity of the theoretical background of the paradigm and associated hypotheses on which to base the experimental design. An exception is the specific case of “sudden degradation of automation”, which seems to fulfill this criterion. Beyond this case, the ability to investigate the AS phenomenon suffers from a “chicken or the egg”-causality dilemma whose circle dependency involves the lack of observational data, the immature theoretic framework on AS and inability to verify causality between factors and AS.

6.1.4 Complexity

The causation of AS may often be explained by an increase of complexity of the socio-technical system. Hollnagel describes “mathematical complexity as a measure of the number of possible states a system can take on, when there are too many elements and relationships to be understood in simple analytic or logical ways.” [46]. An increasing number of states is indeed an inevitable side-effect of introducing automation, where the state of automation can be described by a vector of variables (or internal state vector) representing different modes. In terms of complexity, a significant contribution is made by the interaction loop between humans and automation. The interaction loop brings into play numerous new relationships between the internal states of automation, as well as associated functions, and the states of the operator’s situational awareness, including response actions. Since each new state can interact with other states, this inevitably leads to a multiplication of combinations capable of generating emerging states in the operator-automation interaction. Because of the variability of human performance as well as unforeseeable disturbing events (e.g. automation failure), states may exceed or violate the safety margins established. Faulty states, such as incomplete situation awareness or component failure, propagate more intensely in a highly connected network of interactions, referred to as dysfunctional interaction [24]. The complexity resulting from this promotes not only the propagation of failures but also the range of combinations that could cause an accident. This could be interpreted as a diversification of combinations that may cause accidents. Any combination may show each a lower or even higher probability of occurrence but are at the same time harder to identify and predict from scratch and, consequently, more surprising.

6.1.5 Review Safety Assessment Methods

Previous and current safety assessment methods have generally a Safety-I perspective, where risk analysis has been performed based on historical data. However, Safety-I assessment methods neither offer explicit support to identify emergent states in the human-automation collaboration nor to mitigate causes of AS. There were two exceptions that picked up concepts of how to assess human error probability. First are the methods related to human reliability analysis. They classify and estimate the probability of human error and consider the human as a machine component, following the example of a Probability Risk Assessment (PRA). Aspects of AS are partly covered by "omission of actions" or "error of commission". Secondly, STPA offers the possibility to consider human-automation collaboration and to find emerging states in the control loop.
FRAM is a supplementary framework that can be used in the scope of a safety assessment. It is capable of developing an understanding of how a sociotechnical system works. FRAM can be utilized to model any kind of performance or activity and can therefore be used to develop a model of a system’s functions as a basis for analysis. Consequently, it should be possible to use FRAM to model the effects of automation surprise.
According to regulation (EU) 373/2017, the consequences of changes in a functional system should be expressed in terms of harmful effects of the change and the hazards associated with safety risks. This means that automation surprise can be handled as a failure mode, a starting point for an unwanted event with a harmful effect. Regulation (EU) 373/2017 demands that hazard identification aim to complete coverage of any condition, event, or circumstance related to the change, which could induce a harmful effect individually or in combination. Hence, according to (EU) 373/2017, the hazard

identification process potentially can solely identify automation surprises with a path to a harmful effect.
The methods share common assumptions concerning risk modeling. Those that we see clearly here are linear relationships and binary-distributed event occurrences. Linear models have the advantage of being easily understood and applied by safety analysts. They base on the sequential principle of A causing B causing C and so on. This principle has its roots in the domino model according to Heinrich [50] and allows the analyst to suggest causal relationships in the form of event tree and fault tree models into which safety analysts can embed any event of interest. The application of linear relation comes along with two assumptions. Firstly, there is a well-defined hazard event, which occurs or not (binary distributed). Further, a hazard event is specified using criteria and conditions of occurrence for the purpose to design worst credible case scenarios. The divergence between automation behavior and operator expectation may represent such a hazard that could impact safety negatively. Discrepancies that may arise when using AS phenomena as a hazard may include the following:

- AS is not a clear measurable event with sharp bounds because it is not necessarily related to a certain action. Rather it arises from an invalid expectation of automation behavior, which is simply tied to cognitive processes and ties up the operator's capability to anticipate. The concept of situation awareness might be a good approach to explain AS by a divergence of the anticipation of the situation.
- A closely related side effect of this sequential principle of linear models concerns the nature of hazardous events, which are considered binary distributed events: "On" or "Off". This is at odds with the states and processes of the operator's situational awareness, which are inherently viewed as non-binary but continuous: "More" or "Less".
- As AS might trigger uncountable variants of reactions of the operator, a broad range of possible follow-up scenarios need to be considered in order to assess the consequence of AS.

The AS phenomenon is a generic event, based on retrospective investigation. As such, it has not undergone concretization and contextualization at the application of future implementation, nor does it describe an event with sharp boundaries. It is therefore more the task of the safety analyst to concretize AS into the application context and define what exactly the surprise might be.
Second, there are also preceding and subsequent events connected to the hazard event and depend on conditions or transfer probabilities. Effects and causes shall be assessed, monitored (or observed) and mitigated if tolerable limits are exceeded. Complexity, on the other hand, acts in a network of dependencies that does not feature linearity. This sequential principle is in conflict with the nature of dynamic systems and complex socio-technical systems, showing a large number of interdependencies. Most high-reliability organizations involve a surge in complexity and, consequently, requirements on performance variability. These complex systems are elaborate and contain many details, and the principles of some of the system functions can partly be unknown. These systems are interdependent on other actors in the system and at the system boundaries, and the system changes before the description of the system can be completed.
Linearity of event occurrence does not capture humans as a multidimensional complex of states, processes, and systems knowledge. The inevitable consequence is that relations between states are ignored, which thus contributes to enlarging the gaps in the barriers. If we take the safety assessment at the remote tower as an example, the impact of, for example, a black screen on the visualization depends on the traffic situation and the ATCO's situational awareness. The traffic situation is then dependent on the traffic but also the ATCO clearances, which is based on procedures and ATCO training and experience. Also, the outcome of the situation is dependent on the ATCO and the situation awareness when the failure occurred. This sequence of events is difficult to describe as cause and effect. One could say that there are more conditioning events than basic failure events. In these types of systems, the key to successful performance lies in the ability of the human operators to compensate for incomplete procedures and instructions and adjust their performance accordingly. Hence, adjusting performance is necessary to match the ever-changing system demands, resources, and constraints.

7. Conclusion and Outlook

The study aimed to investigate the reasons why safety assessment methods struggle to proactively identify and mitigate the AS phenomenon, which is, however, essential when climbing the ladder of automation levels. A literature review was chosen that aimed at exploring state-of-the-art findings and knowledge about the factors contributing to AS occurrence. The literature investigation is preliminary and considered a smaller number of references than would have been possible in a full systematic literature search. The study focused on research conducted in the area of cockpit safety because there is a long history of safety incidents related to automation and therefore a larger number of studies have been conducted. Secondly, the goal was to review prospective safety assessment methods to what extent they support the identification and mitigation of AS. A comparison of the AS-contributing factors and assumptions required for safety assessment led to insights into the reasons for the difficulties faced by the methods.

- A list of AS-contributing factors was compiled from the literature reviewed which represents a cross intersection of those lined out in the respective studies.
- The list of factors contributing to AS was classified using the categories and elements as specified in the SOAM standard, a method for investigating incidents and accidents. The review revealed a heterogeneous picture of distribution across categories and elements, indicating a diversity of factors contributing to AS. AS has many forms of manifestation.
- The relationship between complexity and diversity was discussed and rationalized: The more complex, the more diverse the manifestations of AS.
- 17 safety assessment methods were reviewed concerning the support to identify and mitigate AS and related factors. The result is that there are two methods, HRA and STPA, that explicitly support the safety analyst to involve the human contribution to the safety assessment, involving aspects of AS implicitly. HRA supports using predefined human error classification and STPA supports by involving a model of the human-control-loop.
- Safety assessment methods require the specification of hazard events that are well defined in terms of the conditions and limits of their occurrence. AS can occur without any explicit visible action of the human operator. This is because AS occurs at the level of the operator's understanding and anticipation of the situation, a purely cognitive process.
- Complexity is poorly covered using models based on linear relationships because complex systems rely on multi dependencies in a network.
- The binary distribution of event occurrence was identified as a limiting factor, hindering to find suitable descriptions of human behavior that could be defined as a hazard.

In summary, the bottleneck that safety assessment methods must overcome is the fundamental way in which a phenomenon is studied and applied to an appropriate context of any innovative automation that is to be put into operation. AS-related research is based on retrospective analysis whereas safety assessment is a predictive approach. To illustrate this contrast more generally, retrospective analysis provides factors that were identified or observed to be outside predefined tolerances ("what was significant during the occurrence?"). On the other hand, prospective risk analysis is a model-driven approach that identifies hazards and uses causal relationships to assess risk ("What will go wrong from an undesired event?"). The translation between both worlds requires a mature framework that explains and describes AS at the level of application or shows at least how to contextualize.
Further research activities will be based on the following solution approaches

- Using human-in-the-loop simulation techniques and non-interventional measures to identify AS and contributing factors at a proposed socio-technical system intended for operational launch.
- A general understanding of automation states and their interrelationships could increase ATCOs' acceptance and prevent both overconfidence and underconfidence. ATCOs trained in

operating principles and automation implementation could learn to act at a higher level of understanding and avoid surprises by focusing their attention on known and potential automation problems. This should increase awareness of automation principles, which can be referred to as automation awareness, building on mode-of-operation awareness.

- In the context of operator awareness of automation, developing use cases to implement automation rules and pre-programmed logic is key to avoiding AS. Analyzing and coding work-as-done in standard annotation forms would add validity to the design process with respect to the cognitive processes that operators go through. The design process could better focus automation on specific parts of the work that are to be automated. Automation is then a more consistent part of the workflow and provides more on-demand support to the operator. This may add more "operator awareness" to the automation.

References

[1] Single European Sky ATM Research 3 Joint Undertaking. *European ATM Master Plan - Executive view*. 2020th ed. SESAR Joint Undertaking. 2020.

[2] Eurocontrol. *Specification for Advanced-Surface Movement Guidance and Control System (A-SMGCS) Services*. 2nd ed. EUROCONTROL-SPEC-171. Eurocontrol. 2020. ISBN: 978-2-87497-097-9.

[3] EASA. *Annual Safety Review 2021*. EASA, 2021.

[4] Jonas Lundberg, Rogier Woltjer, and Billy Josefsson. "A method to identify investigative blind spots (MIBS): Addressing blunt-end factors of ultra-safe organizations' investigation-work-as-done". In: *Safety Science* 154 (2022), p. 105825. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2022.105825. URL: https://www.sciencedirect.com/science/article/pii/S0925753522001643.

[5] Everett Palmer. "Oops, it didn't arm-a case study of two automation surprises". In: *Proceedings of the Eighth International Symposium on Aviation Psychology*. Ohio State University Columbus, Ohio. 1995, pp. 227–232.

[6] Wayne L Martin, Patrick S Murray, and Paul R Bates. "The effects of startle on pilots during critical events: A case study analysis". In: *Griffith Univ. Aerosp. Strateg. Study Cent* (2012), pp. 387–394.

[7] Javier Rivera et al. "Startle and surprise on the flight deck: Similarities, differences, and prevalence". In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 58. 1. SAGE Publications Sage CA: Los Angeles, CA. 2014, pp. 1047–1051.

[8] S. Dekker and Lunds universitet. School of Aviation. *Report of the Flight Crew Human Factors Investigation Conducted for the Dutch Safety Board Into the Accident of TK1951, Boeing 737-800 Near Amsterdam Schiphol Airport, February 25, 2009*. Lund University, School of Aviation, 2009.

[9] J.N. Field et al. *Startle Effect Management*. NLR-CR-2018-242. NLR, 2021.

[10] Tor Erik Evjemo. "Sikkerhet og autonomi i norsk luftfart utfordringer og muligheter". In: *SINTEF rapport* (2018).

[11] Sidney Dekker. *The field guide to human error investigations*. Routledge, 2017.

[12] K Abbott, David McKenney, and Paul Railsback. "Operational use of flight path management systems–Final report of the Performance-based operations Aviation Rulemaking Committee". In: *Commercial Aviation Safety Team Flight Deck Automation Working Group* (2013).

[13] Dong-Han Ham. "Safety-II and Resilience Engineering in a Nutshell: An Introductory Guide to Their Concepts and Methods". In: *Safety and Health at Work* 12.1 (2021), pp. 10–19. ISSN: 2093-7911. DOI: https://doi.org/10.1016/j.shaw.2020.11.004. URL: https://www.sciencedirect.com/science/article/pii/S2093791120303619.

[14] Robert De Boer and Sidney Dekker. “Models of Automation Surprise: Results of a Field Survey in Aviation”. In: *Safety* 3.3 (2017). ISSN: 2313-576X. DOI: 10.3390/safety3030020. URL: https://www.mdpi.com/2313-576X/3/3/20.

[15] Nadine B Sarter, David D Woods, Charles E Billings, et al. “Automation Surprises”. In: *Handbook of human factors and ergonomics* 2 (1997), pp. 1926–1943.

[16] Nadine B Sarter and David D Woods. “How in the world did we ever get into that mode? Mode error and awareness in supervisory control”. In: *Human factors* 37.1 (1995), pp. 5–19.

[17] N. McDonald. “Do I trust thee? An approach to understanding trust in the domain of air traffic control”. English. In: *IET Conference Proceedings* (Jan. 2001), 104–109(5). URL: https://digital-library.theiet.org/content/conferences/10.1049/cp_20010441.

[18] Sidney Dekker and Erik Hollnagel. “Human factors and folk models”. In: *Cognition, Technology & Work* 6.2 (2004), pp. 79–86.

[19] Mica R Endsley. “Toward a theory of situation awareness in dynamic systems”. In: *Situational awareness*. Routledge, 2017, pp. 9–42.

[20] David B. Kaber and Mica R. Endsley. “Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety”. In: *Process Safety Progress* 16.3 (1997), pp. 126–131. DOI: https://doi.org/10.1002/prs.680160304. eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/prs.680160304. URL: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/prs.680160304.

[21] Barry Strauch. “Ironies of Automation: Still Unresolved After All These Years”. In: *IEEE Transactions on Human-Machine Systems* 48.5 (2018), pp. 419–433. DOI: 10.1109/THMS.2017.2732506.

[22] Jonas Lundberg and Billy Josefsson. “A Pragmatic Approach to Uncover Blind Spots in Accident Investigation in Ultra-safe Organizations - A Case Study from Air Traffic Management”. In: *Advances in Human Error, Reliability, Resilience, and Performance*. Ed. by Ronald L. Boring. Cham: Springer International Publishing, 2019, pp. 199–210. ISBN: 978-3-319-94391-6.

[23] Erik Hollnagel. *Safety–I and Safety–II: The Past and Future of Safety Management (1st ed.)* CRC press, 2014. URL: https://doi.org/10.1201/9781315607511.

[24] Nancy Leveson. “A new accident model for engineering safer systems”. In: *Safety Science* 42.4 (2004), pp. 237–270. ISSN: 0925-7535. DOI: https://doi.org/10.1016/S0925-7535(03)00047-X. URL: https://www.sciencedirect.com/science/article/pii/S092575350300047X.

[25] Peter Brooker. “SESAR safety decision-making: Lessons from environmental, nuclear and defense modeling”. In: *Safety Science* 48.7 (2010), pp. 831–844. ISSN: 0925-7535. DOI: https://doi.org/10.1016/j.ssci.2010.03.015. URL: https://www.sciencedirect.com/science/article/pii/S0925753510000871.

[26] Erik Hollnagel and David D Woods. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering (1st ed.)* CRC press, 2005. URL: https://doi.org/10.1201/9781420038194.

[27] Erik Hollnagel and David D. Woods. “Cognitive Systems Engineering: New wine in new bottles”. In: *International Journal of Man-Machine Studies* 18.6 (1983), pp. 583–600. ISSN: 0020-7373. DOI: https://doi.org/10.1016/S0020-7373(83)80034-0. URL: https://www.sciencedirect.com/science/article/pii/S0020737383800340.

[28] M.R. Endsley. *Testimony to the United States House of Representatives: Hearing on Boeing 737-Max8 Crashes*. Tech. rep. Human Factors and Ergonomics Society, Dec. 2019.

[29] Eurocontrol. *EAM2/GUI8 Guidelines on the Systematic Occurrence Analysis Methodology (SOAM)*. 1.0. Eurocontrol. 2005.

[30] Lance Sherry and Robert Mauro. “Controlled Flight into Stall (CFIS): Functional complexity failures and automation surprises”. In: *2014 Integrated Communications, Navigation and Surveillance Conference (ICNS) Conference Proceedings*. 2014, pp. D1-1-D1–11. DOI: 10.1109/ICNSurv.2014.6819980.

[31] Frederic Dehais et al. “"Automation Surprise" in Aviation: Real-Time Solutions”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI ’15. Seoul, Republic of Korea: Association for Computing Machinery, 2015, pp. 2525–2534. ISBN: 9781450331456. DOI: 10.1145/2702123.2702521. URL: https://doi.org/10.1145/2702123.2702521.

[32] Amy Rankin, Rogier Woltjer, and Joris Field. “Sensemaking following surprise in the cockpit—a re-framing problem”. In: *Cognition, Technology & Work* 18.4 (2016), pp. 623–642.

[33] Robert de Boer and Karel Hurts. “Automation Surprise: Results of a Field Survey of Dutch Pilots”. In: *Aviation Psychology and Applied Human Factors* 7 (Apr. 2017), pp. 28–41. DOI: 10.1027/2192-0923/a000113.

[34] Raja Parasuraman and Dietrich H. Manzey. “Complacency and Bias in Human Use of Automation: An Attentional Integration”. In: *Human Factors* 52.3 (2010). PMID: 21077562, pp. 381–410. DOI: 10.1177/0018720810376055. eprint: https://doi.org/10.1177/0018720810376055. URL: https://doi.org/10.1177/0018720810376055.

[35] Mika Endsley. “Situation Awareness in Aviation Systems”. In: *Handbook of aviation human factors: Second edition*. Ed. by J.A. Wise, V.D. Hopkin, and D.J. Garland. 2nd. CRC Press, 2016. Chap. 12, pp. 268–289.

[36] Nadine B. Sarter and David D. Woods. “Team Play with a Powerful and Independent Agent: Operational Experiences and Automation Surprises on the Airbus A-320”. In: *Human Factors* 39.4 (1997). PMID: 11536850, pp. 553–569. DOI: 10.1518/001872097778667997. eprint: https://doi.org/10.1518/001872097778667997. URL: https://doi.org/10.1518/001872097778667997.

[37] David D Woods and Nadine B Sarter. “Learning from automation surprises and going sour accidents”. In: *Cognitive engineering in the aviation domain* (2000), pp. 327–353.

[38] Marvin Rausand and Stein Haugen. *Risk assessment: theory, methods, and applications.* John Wiley & Sons, Ltd, 2020. ISBN: 9781118281116. DOI: https://doi.org/10.1002/9781118281116. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118281116. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118281116.

[39] M.H.C. Everdij and H.A.P. Blom. *Safety Methods Database*. Version 1.2. Available at. Netherlands Aerospace Centre NLR, 2020. URL: http://www.nlr.nl/documents/flyers/SATdb.pdf.

[40] B. Kirwan. *A Guide to Practical Human Reliability Assessment ()*. 1st ed. CRC Press, 1994. URL: https://doi.org/10.1201/9781315136349.

[41] Thomas Fylking and Christian Bjursten Carlsson. *Skyddsbarriäranalys (LOPA): Vägledning för val av numeriska data*. 1st ed. IPS, 2016. URL: http://libris.kb.se/bib/19704582.

[42] Eurocontrol. *Safety Assessment Methodology - A framework of methods and techniques to develop safety assessments of changes to functional systems*. Version 2.1. Eurocontrol. 2006. URL: https://www.eurocontrol.int/tool/safety-assessment-methodology.

[43] Nancy G Leveson. *Engineering a safer world: Systems thinking applied to safety*. 2016. URL: http://library.oapen.org/handle/20.500.12657/26043.

[44] Leonid Sedov et al. “Qualitative and Quantitative Risk Assessment of Urban Airspace Operations”. In: *11th SESAR Innovation Days,* SESAR Joint Undertaking. SESAR Joint Undertaking, 2021.

[45] European Commission. “Commission Implementing Regulation (EU) 2017/373 of 1 March 2017 laying down common requirements for providers of air traffic management/air navigation services and other air traffic management network functions and their oversight, repealing Regulation (EC) No 482/2008, Implementing Regulations (EU) No 1034/2011, (EU) No 1035/2011 and (EU) 2016/1377 and amending Regulation (EU) No 677/2011 (Text with EEA relevance.)” In: *Official Journal of the European Union* (2017).

[46] Erik Hollnagel. *FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*. Oct. 2012. ISBN: 978-1-4094-4551-7. DOI: 10.1201/9781315255071.

[47] Sidney Dekker. *Report of the flight crew human factors investigation conducted for the Dutch safety board into the accident of TK1951, Boeing 737-800 near Amsterdam Schiphol Airport, February 25, 2009*. Lund University, School of Aviation, 2009.

[48] KNKT. *Aircraft Accident Investigation Report. PT. Lion Airlines Boeing 737 (MAX); PK-LQP Tanjung Karawang, West Java, Republic of Indonesia 29 October 2018*. Tech. rep. Komite Nasional Keselamatan Transportasi (KNKT), 2018.

[49] Addis Abada. *Interim Investigation Report on Accident to the B737-8 (MAX) Registered ET-AVJ operated by Ethiopian Airlines*. Tech. rep. Ethiopian Civil Aviation Authority, Ministry of Transport, 2020.

[50] Herbert William Heinrich et al. “Industrial Accident Prevention. A Scientific Approach.” In: *Industrial Accident Prevention. A Scientific Approach.* Second Edition (1941).

Author Biography

Dr. Lothar Meyer is a safety engineer at the Air Navigation Services of Sweden LFV. He holds a degree in electrical engineering and a doctorate in air traffic services from the Technische Universität Dresden, where he worked as a research associate in the field of aviation safety. His areas of expertise covers socio-technical systems, risk assessment and airport surveillance technologies

Mr. Christian Bjursten Carlsson is a senior principal human factors and safety consultant for Tapora engaged in various ATM projects for the Air Navigation Services of Sweden LFV. Christian has a master’s degree in chemical engineering from Lund University, Sweden, supplemented with studies in psychology, human factors, and system safety at Lund University. In addition, he holds a commercial pilot license, a flight instructor rating, and has experience as airline pilot on the Boeing 737 NG. Christian has extensive international experience from projects in the aviation industry, air navigation service providers, chemical process industry, medical technology, nuclear power plants, and oil and gas operators. Christian’s main focus areas are human factors, system safety, and risk management in safety-critical industries.

Dr. Åsa Svensson is a human factors expert at the Air Navigation Services of Sweden LFV. She holds a PhD in design focusing on human-automation collaboration from Linköping University, Sweden. She currently works with human factors questions related to automation and situation awareness within air traffic control.

Mr. Maximilian Peukert received a MSc degree in psychology and human performance in 2017. After working as a research associate and lecturer at the chair of engineering psychology at Technische Universität Dresden, he is since 2018 active as a Human Factors Specialist at the Air Navigation Services of Sweden LFV.

Mr. Lars Danielson is a safety engineer at Saab Digital Air Traffic Solutions. He holds a Master of science in Industrial Engineering and Management from Linköping University, Sweden. He currently works with safety assessments related to remote air traffic control.

Mr. Billy Josefsson is a senior Air Traffic Controller with a background in psychology. Since 1994 active within research and development in human performance, safety and human factors worldwide. Since 2014 he is Manager for Automation and Human Performance at the Air Navigation Services of Sweden LFV.

Copyright Statement