Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots' behavior

Patrick Gontar[a,*], Hendrik Homans[a], Michelle Rostalski[a], Julia Behrend[b], Frédéric Dehais[b], Klaus Bengler[a]

[a] Chair of Ergonomics, Technical University of Munich, Munich, Germany
[b] Institut Supérieur de l'Aéronautique et de l'Espace, Université Fédérale de Toulouse Midi-Pyrénées, Toulouse, France

Keywords:
Cyber security
Human factors
Simulator study
Pilots
Trust
Workload
Eye-tracking

A B S T R A C T

The increasing prevalence of technology in modern airliners brings not just advantages, but also the potential for cyber threats. Fortunately, there have been no significant attacks on civil aircraft to date, which allows the handling of these emerging threats to be approached proactively. Although an ample body of research into technical defense strategies exists, current research neglects to take the human operator into account. In this study, we present an exploratory experiment focusing on pilots confronted with a cyber-attack. Results show that the occurrence of an attack affects all dependent variables: pilots' workload, trust, eye-movements, and behavior. Pilots experiencing an attack report heavier workload and weakened trust in the system than pilots whose aircraft is not under attack. Further, pilots who experienced an attack monitored basic flying instruments less and their performance deteriorated. A warning about a potential attack seems to moderate several of those effects. Our analysis prompts us to recommend incorporating cyber-awareness into pilots' recurrent training; we also argue that one has to consider all affected personnel when designing such training. Future research should target the development of appropriate procedures and training techniques to prepare pilots to correctly identify and respond to cyber-attacks.

1. Introduction

The exponentially increasing incidence of cyber-attacks is a growing problem in various private and public domains (Wilshusen, 2013). These range from personal cell phones and computers to critical infrastructures—including that of civil aviation (Elias, 2015; Zan, d'Amore and Di Camillo, 2016). A cyber-attack implies deliberate actions "to alter, disrupt, deceive, degrade, or destroy adversary computer systems or networks or the information and (or) programs resident in or transiting these systems or networks" (Owens et al., 2009, p. S-1). In aviation, the use of complex computer information technology such as that for fly-by-wire or flight management systems has intensified in recent decades. This trend has created potential vectors for cyber-attacks (Sampigethaya and Poovendran, 2013). The interdependence between complex aircraft systems and their integration into a modern airliner can easily propagate the effects of a cyber-attack from one system to another (Haass et al., 2016). Vereinigung Cockpit (2017) gave an overview of how interlinked the different systems in the aviation domain are, and where possible attack vectors might exist (see Fig. 1).

Several national and international aviation agencies (e.g., American Institute of Aeronautics and Astronautics, 2013; European Aviation Safety Agency, 2016; Iasiello, 2014; International Civil Aviation Organization, 2012, 2016; International Federation of Air Line Pilots' Associations, 2013; Lim, 2014) have already acknowledged that the civil aviation domain is potentially subject to cyber-attacks. Cyber-attacks against aircraft are still extremely rare at the time of writing; however, their increasing incidence in the future is highly probable and may lead to catastrophes, especially given the current rate of development in information technologies (International Civil Aviation Organization, 2016). Fox (2016) points out that although nothing serious has happened so far, it is a question of *when* rather than *if*. The vulnerability of commercial aircraft systems was highlighted by the U.S Department of Homeland Security, which was able to penetrate a commercial aircraft via radio frequency communication in 2016 (Biesecker, 2017). The airline industry as well as regulators take this problem very seriously and are following different approaches (American Institute of Aeronautics and Astronautics, 2013; Iasiello, 2014) and also amended regulations (Federal Aviation Administration, 2013; 2014) to try to prevent potential attacks.

However, these approaches focus mainly on technical means to

* Corresponding author. Chair of Ergonomics, Technical University of Munich, Boltzmannstr. 15, 85748, Garching, Germany.
E-mail address: gontar@tum.de (P. Gontar).

https://doi.org/10.1016/j.jairtraman.2018.01.004

Fig. 1. Visualization of interlinked systems in the civil aviation domain showing several potential vectors for cyber-attacks. The figure is based on Vereinigung Cockpit (2017).

render potential attacks technically improbable. Very recently, a patent was granted to Boeing, in which the inventors suggest a new system to evaluate pilots' response to cyber-attacks in a simulation environment (Nguyen et al., 2017). In their description of the evaluation system, Nguyen et al. (2017) argue that "… because the pilot is such an integral part of the operation and control of the aircraft, pilot reaction to a cyber-attack is important" (p. 7). Besides erecting formidable technical and organizational barriers to eliminate security hazards before they reach the cockpit, we agree with Nguyen et al. (2017) that the human operator has to be integrated as a defense layer (Boyce et al., 2011; Langton and Baker, 2013), if not as the last line of defense (Vereinigung Cockpit, 2017). In this context it is important to distinguish between *safety and security*. Pièrre-Cambacédès and Chaudet (2010) analyzed the usage and definition of both constructs extensively. They found not only that several researchers fail to explicitly define what they mean by one or the other, but also that very different definitions are used in different domains. Coming from the human factors domain, we favor the distinction from Firesmith (2003, p. 2) who defines safety as "the degree to which *accidental* harm is prevented, detected, and properly reacted to." Common safety issues might arise in the context of fatigue (Caldwell, 2012; Rosekind et al., 1994), loss of manual flying skills (Haslbeck and Hoermann, 2016), complex task switching (Gontar et al., 2017a,b), or technical malfunctions involving effortful problem solving and decision-making (Mosier and Fischer, 2010; Orasanu and Fischer, 2014) as well as intense team-communication (Gontar et al., 2017a,b). Firesmith (2003, p.14) defines security as "the degree to which *malicious* harm to a valuable asset is prevented, detected, and reacted to." Security is often seen as a technical challenge, although a successful cyber breach could evoke pilot reactions resembling those to a safety problem. However, from a human factors perspective, we think that pilots perceive differences between a cyber-breach-induced malfunction and a purely technical one. We point out these differences in the next section.

1.1. A human factors approach to cyber-attacks

Unfortunately, researchers have neglected human operators' response behavior in earlier cyber security research (Mancuso et al., 2014; Proctor and Chen, 2015). Horowitz and Lucero (2016) and Heiges et al. (2015) used a scenario with a manipulated navigation system showing false waypoints. Their main interest, however, was in investigating which security requirements would be usefully satisfied in engineering tools. Human factors analysis showed pilots' explicit wish for technical support during a cyber-attack as well as their concern about making ill-informed decisions. The major issue is that a successful attack exposes pilots to great uncertainty (Dutt et al., 2013; Hirshfield et al., 2015). Although individual cues might be ambiguous during the very infrequent occasions when technical malfunctions occur, pilots can normally apply procedures to solve the associated technical problems. Faced with a purely technical problem, pilots can also anticipate not only their own course of action but also the aircraft's behavior when, for example, a hydraulic system is leaking. The pilots know that they will receive an alert that hydraulic pressure is too low, maybe followed by another alert that the fluid level is also too low. Further, the pilots (depending on the aircraft type) might receive information from the aircraft system about how the specific technical malfunction will affect the aircraft's performance. In the example of a hydraulic system burst, the pilots can anticipate how this malfunction will affect their future flight—that their high-lift system will move slower, for instance—so that they can prepare mentally. When pilots face a cyber-attack, in contrast, they do not know whether the cues are trustworthy, or clear cues that an aircraft-borne system is under a cyber-attack might be absent. The pilots know neither whether the problem is solvable with their checklists and procedures, nor whether they will experience another problem right afterward. If pilots automatically follow their procedures in such cases, one can imagine potential attackers exploiting that knowledge to manipulate the pilots' behavior. Handling cyber-attacks, which are likely characterized by ambiguous cues lacking clear response options, becomes very effortful and also difficult for the pilots. The cue-clarity model that Orasanu et al. (1993) developed helps to understand how *cue-clarity* and *response option* availability can affect pilots' decision-making and performance.

1.2. An issue of cue-clarity

Cue-clarity describes a cue's clearness or ambiguity. An example for a clear cue could be a 'green hydraulic low pressure' warning in the abovementioned loss of a hydraulic system, while an ambiguous cue could be something like 'expect weather changes on-route'. Indeed, Dismukes et al. (2007) argue that weather information displays

onboard an aircraft are not specific and accurate enough to allow explicit weather assessment. *Response options* describes the number of prescribed options available in a situation. Orasanu et al. (1993) distinguish three categories: *single* option (only one prescribed response), *multiple* options (several prescribed responses from which the pilots have to choose one), and *none* (no prescribed responses are available and pilots have to generate new ones). Pilots experience extensive workload when confronting a situation in which they have no prescribed response options available (Orasanu et al., 1993). Cue ambiguity renders workload even higher. Both conditions prevail during a successful cyber-attack. Prior research has demonstrated the influence of excessive cognitive demand on decision-making under uncertainty (Heereman and Walla, 2011) as well as the deterioration of decision quality due to perceived stress (Starcke and Brand, 2016). Furthermore, interpreting ambiguous cues engenders cognitive conflicts known to impair pilots' performance (Dehais et al., 2003; Dehais et al., 2010). Situations featuring unclear cues monopolize human operators' conflict resolution resources to the detriment of monitoring primary parameters (Dehais et al., 2015; Tessier and Dehais, 2012) which can in turn reduce performance.

As cyber-attacks are not currently a topic in airline training, individual pilots will have to rely on their own knowledge and personal experience with ordinary malfunctions to generate response options for detecting a cyber-attack or making sense of available information (Dutt et al., 2013; Grazioli, 2004; Hirshfield et al., 2015). Doing so could become dangerous as pilots' assumption might not hold true during a cyber-attack. As pilots undergo initial and mandatory recurrent training to handle known safety hazards, one might assume that they are also experts in handling situations such as cyber-attacks. To gain sufficient expertise in handling such unforeseen events, pilots need to 1) have the chance to learn and experience the relevant cues, and 2) the cues have to be valid (Klein, 1999). Fortunately, opportunities to learn relevant cues arise very infrequently during operation, because pilots very seldom experience severe technical malfunctions and have not yet experienced real cyber breaches during normal operation. This means that if pilots are to become experts at handling unusual, ambiguous situations, then pilot training has to provide an appropriate environment for it. Current training, during which pilots often know what is going to happen and can mentally prepare for it, seems not to feature sufficiently valid cues when it comes to handling unforeseen events (Bergström et al., 2014; Casner et al., 2013; Dahlström et al., 2009). Considering the ambiguous cues in cyber-attacks as opposed to ordinary technical malfunctions, it becomes obvious that they severely hinder even experts from deciding quickly and correctly (see Orasanu et al., 1993).

1.3. Deception detection model

Given the lack of possibilities to gain sufficient expertise in handling cyber-attacks and the fact that pilots cannot rely on their experience in handling them, questions remain concerning how pilots detect and handle cyber-attacks. Grazioli's deception detection model (2004) focuses on individuals' information processing and on detecting cyber-attacks or deceptive information on the Internet. His model posits four processes involved in detecting cyber-attacks. The first, *activation,* assumes that individuals continually compare information (cues) from the environment with their experience of and expectations about a situation. For instance, pilots compare their vertical to their horizontal speed during an approach. Doing so enables them to calculate the descent angle to ensure not capturing one of the instrument landing system' side lobes. Individuals become cognitively activated when a cue contradicts their expectations about a situation. This would be the case when the indicated airspeed and the vertical speed do not fit the desired glide path angle. In other words, their attention is drawn to the discrepancies between the cues and their expectations. Individuals *generate* different *hypotheses* to explain these discrepancies. In the deception detection model's next process, every *hypothesis* is *evaluated* by comparing it to a domain-specific criterion, which may be difficult to define depending on the context and the individual's knowledge and expertise. Users accept or reject a hypothesis based on how it compares to the criterion. All of the accepted hypotheses are ultimately combined into a *global assessment* based upon which the user decides whether or not the information is deceptive. This theory holds that hypothesis generation taking a cyber-attack into account is the key element. Only when generating a correct hypothesis regarding a potential cyber-attack would pilots be able to detect one, decide correctly, and act accordingly. Hypothesis generation here depends mainly on cue interpretations and personal experience. Although the deception detection model explains how operators decide whether information is deceptive or trustworthy, it doesn't predict their potential reaction to it. The research presented in this paper aims to fill this gap by analyzing how pilots react to a successful cyber-attack and how such an attack influences their behavior.

1.4. Research questions

Deception detection theory (Grazioli, 2004) requires the operator to recognize a cyber-attack by following an analytic procedure to generate and evaluate hypotheses. Different models and views all suggest that such an analytic procedure places severe demands on cognitive resources (Bobko et al., 2014; Boyce et al., 2011; Dutt et al., 2013; Grazioli, 2004; Hirshfield et al., 2015; Rasmussen, 1983). Further, Orasanu et al. (1993) argue that ambiguous cues such as those, we argue, present during a cyber-attack also increase workload. Results of empirical research indicate a positive relation between the occurrence of a cyber-attack and operators' workload (Hirshfield et al., 2015). We thus formulate the following research questions:

RQ 1: Does a cyber-attack relate positively to pilots' workload?
RQ 2: Does a cyber-attack warning attenuate the attack's effect on pilots' workload?

From an operator's perspective, a cyber-attack can be interpreted as a system error or as an unreliable part of a system. Furthermore, cue ambiguity plays an important role as the pilots know neither whether the cues are reliable nor what will happen next. Experiments in several domains have shown that a system's high reliability and low error rate strengthen trust in the system and vice versa (Dzindolet et al., 2003; Moray et al., 2000; Vries et al., 2003; Yeh and Wickens, 2001). Weakened operator trust can lead to system disuse (Muir, 1987) or non-respondence, especially when the unreliability of warning systems results in false alarms or misses (Manzey et al., 2014). Experiencing false alarms can lead to an operator's reluctance to acknowledge true alarm as well as to delayed responses to alarms—an effect often referred to as the cry-wolf effect (Breznitz, 1984; Wickens et al., 2009), which is known to intensify under high workload conditions (Bliss and Dunn, 2000). This phenomenon can also spread to other functions (Lee and Moray, 1992), although other empirical studies (e.g., Bahner et al., 2008) found contradicting results. Furthermore, behavior predictability influences trust (Lee and See, 2004; Rempel et al., 1985), which may weaken when discrepancies between expected and actual system behavior are perceived as may happen due to a cyber-attack.

RQ 3: Does a cyber-attack influence pilots' trust in the system?
RQ 4: Does a cyber-attack warning attenuate the attack's effect on pilots' trust?

The ambiguity of the cues associated with lacking response options in face of a cyber-attack require pilots to apply problem-solving strategies. Previous research has suggested that eye movements could help to better understand pilots' behavior and information acquisition strategies on the flight deck (Dehais et al., 2008; Gontar and Mulligan, 2015, 2016; Haslbeck et al., 2012; Haslbeck and Zhang, 2017;

Lefrançois et al., 2016; Reynal et al., 2016; Sarter, Mumaw and Wickens, 2007). For instance, some studies have shown that conflicting and unexpected situations impair visual scanning leading to either attentional tunneling (Régis et al., 2014) or excessive and inefficient visual search patterns (Dehais et al., 2015). Further, Hergeth et al. (2016) have shown that decreased monitoring frequency is associated with strengthened trust in the system, which itself is associated with system reliability (see also Oakley et al., 2003). We hypothesize that pilots' information acquisition behavior changes when a cyber-attack or warning confronts them.

RQ 5: Does a cyber-attack influence pilots' visual information acquisition strategies?
RQ 6: Does a cyber-attack warning attenuate the attack's effect on pilots' visual information acquisition strategies?

Since pilots are currently not trained in how to react to a cyber-attack or how to handle a respective warning, we expect a wide variety of responses from them as they have to develop new response options (see Orasanu et al., 1993). The first question, however, is whether or not the pilots' performance deteriorates under cyber-attack.

RQ 7: Does a cyber-attack influence pilots' performance?
RQ 8: Does a cyber-attack warning attenuate the attack's effect on pilots' performance?

The extent to which pilots react to a cyber-attack is not yet sufficiently investigated. This paper does not suggest a holistic solution to handling cyber-attacks within the cockpit, but serves as a first step toward understanding pilots' behavior in such situations and suggests directions for future research. This might be the first experimental study investigating the aforementioned aspects of workload, trust, visual information acquisition behavior, and performance when pilots face a cyber-attack in the cockpit.

2. Method

2.1. Test design

To differentiate between the influences of a cyber-attack and a warning about it, the experimental study was conducted as a mixed (2 × 2) design with the within-factor, *warning,* and the between-factor, *attack*. One group experienced a cyber-attack (attack group); one group did not (no-attack group). Participating pilots in both groups were flying the same five scenarios in a flight simulator whereby two trials were experimental trials and the others distractors. In one experimental trial, both groups received a warning from the air traffic controller that their aircraft might have been attacked (warning condition). We deliberately implemented an ambiguous warning in order to induce a high degree of uncertainty.

We use the terminology of the receiver operator characteristic (ROC) as introduced by Youden (1950) to make the study and the results easier to comprehend. For our scenario, the air traffic controller is regarded as the operator who *detects* the attack resulting in a warning; the attack is the *stimulus* (see Table 1). A present stimulus is defined as a *hit* if it is detected and a *miss* if not. An absent stimulus is defined as *false alarm* if it is detected and a *correct rejection* if it is not.

Table 1
Receiver operator characteristic for the groups and conditions.

	Warning	No-warning
Attack	Hit	Miss
No-attack	False alarm	Correct rejection

2.2. Sample

In total, twenty-two male pilots were recruited to participate in the experiment. Each was randomly assigned to one of the experimental groups. The participants' mean age was $M = 38.27$ years ($SD = 11.55$ years) ranging from 25 years to 63 years in the attack group and $M = 28.45$ years ($SD = 5.11$ years) ranging from 22 years to 41 years in the no-attack group. Participants in the attack group had $M = 6062.73$ h ($SD = 7106.38$ h), ranging from 270 h to 23,000 h of flight experience; participants in the no-attack group had $M = 950.86$ h ($SD = 971.62$ h), ranging from 140 h to 3550 h. Participants received no compensation except their travel expenses for their participation; furthermore, all pilots were eligible to voluntarily participate in a lottery for an online-shop voucher.

Minimum participation requirements were defined so that pilots needed to have (1) passed a theoretical exam for an airline transport pilot license (ATPL) or multi-crew pilot license (MPL), (2) an instrument rating with more than 50 flight hours of instrument flight (IFR), and (3) more than 32 flight hours in a turbine airplane.

2.3. Dependent measures

2.3.1. Workload

Pilots' workload was measured using the raw version of the NASA TLX (Byers et al., 1989). The raw version of the NASA TLX (*r*TLX) does not weigh the six different sub-scales (mental demand, physical demand, temporal demand, performance, effort, and frustration) against each other and is therefore easier for the participants to use while still showing "essential equivalence with TLX" (Byers et al., 1989, p. 484). Overall workload is defined as the mean value of the six subscales. It ranges from 0 to 100, where 100 represents the heaviest possible workload.

2.3.2. Trust

Pilots' trust in the system was evaluated with the German version (Gold et al., 2015) of the trust questionnaire originally developed by Jian et al. (2000) and validated by Spain et al. (2008). In the trust questionnaire, participants had to rate their agreement with 12 statements on a seven-point Likert scale from *not at all* to *extremely*. Five statements are formulated in a negative manner and seven are formulated in a positive manner. Overall trust is calculated as the mean of each item on a scale from 1 to 7, where 7 indicates the strongest possible trust.

2.3.3. Gaze behavior

Pilots' gaze behavior was quantified by analyzing the attention ratio toward three defined sets of areas of interest (AoIs) (see Fig. 2). The attention ratio toward a specific AoI set is defined as the duration of all glances onto this specific AoI set divided by the total duration of all glances to any AoI set. For the analysis of the gaze behavior, we focus on the interval that starts when the attack group comes under attack and lasts for 25 s. This allows enough time for the pilot to determine whether or not there is an attack.

We defined the three AoI sets representing the pilot's main tasks: *Aviate* (A), *navigate* (N), and *system management* (M). *Aviate* consisted of the two speed indications on the primary flight display and the navigation display, the altimeter, and the attitude indicator. The *navigate* set comprised glances on both heading indicators, the three distance-measuring equipment systems (DMEs), and the map. The *management* set contained the display showing the system status like an electronic centralized aircraft monitor (ECAM) as well as the engine indicators. The pilots' communicate task which priority would follow the *navigate* task and precede the *management* task is not considered here since we are focusing on gaze behavior only.

Fig. 2. Cockpit interior from participant's point of view showing the definition of AoI sets: *Aviate* (A), *navigate* (N), and *system management* (M).

2.3.4. Performance

Pilots' performance criteria were defined by whether or not they solved the problem correctly. To decide whether a decision was correct or not, flight track parameters were used with a threshold defined according to the approach map. We did not use expert evaluation, because we had learned from previous research that rater reliability is not always sufficient (Gontar and Hoermann, 2015). Furthermore, classical performance measures such as flight-track errors or procedural errors do not reflect performance in the context of the presented scenario—especially since no appropriate checklists exist yet. In addition to the quantifiable dependent measures, we further chose a qualitative approach and asked the pilots (1) what would have helped them in the situation just-experienced and (2) what additional information they would like to have had from the air traffic controller.

2.4. Materials and apparatus

2.4.1. Flight simulator

The experiment took place in a fixed-base simulator, which was equipped with a generic glass cockpit and the flight dynamics of a Dornier 728 (see Haslbeck and Bengler, 2016). The flight dynamic model was chosen to control for the effects of familiarity that some pilots might have with the Airbus A320. The auto-flight system and the auto-throttle were disconnected so that pilots had to fly manually all the time. Manual flight was forced to serve as a secondary task to further increase workload during problem solving.

2.4.2. Eye-tracking system

We used the Dikablis Professional binocular head-mounted eye-tracker from Ergoneers to track the pilots' eye movements. The system samples at 60 Hz and provides a glance direction accuracy of ± 1.8° visual angle. The system had to be calibrated prior to use for every participant to ensure sufficient precision. External markers on the instrument panel in form of barcodes were used to reference the gaze to the respective instruments in the cockpit.

2.4.3. Scenario basis

The scenario started 5 nm out of the Munich (MUN) VOR[1] at an altitude of 5500 ft heading south before getting the clearance direct MUN for the non-standard instrument landing system (ILS) approach to runway 26L (see Fig. 3). Shortly after passing D16MDF (distance of 16 nm from the Milldorf (MDF) VOR), pilots were required to descent according to the approach to a target altitude of 5000 ft. On their way to D3.8MDF, air traffic control contacted the pilots and asked them to confirm their current speed and altitude so the pilots were again prompted to carefully fly according to the requirements. Pilots received the landing clearance when passing D11MDF. When the pilots reached D3.8MDF, they had to turn left to intercept the outbound radial 311° of MDF to follow the approach. The scenario ended on the way to NELBI since the final approach was irrelevant to the experiment.

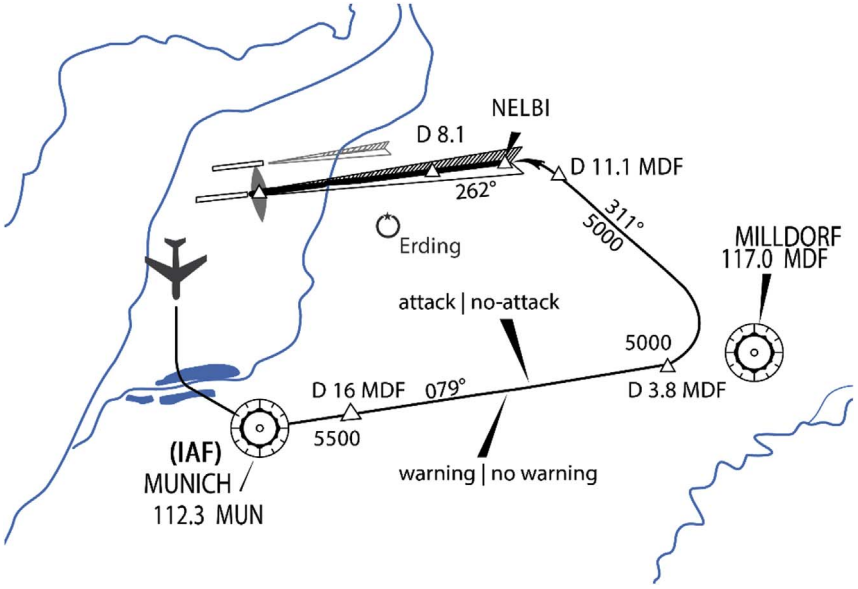

Fig. 3. Scenario including the approach to runway 26L in Munich.
Reduced map excerpt is for the purpose of illustration only. NOT FOR NAVIGATION.

2.4.4. Attack group

For the *attack* group, we decided to simulate manipulation of a navigation aid, namely the MDF VOR station. It is important that this reflects the manipulation of a ground facility rather than of an airborne system in the aircraft. The attack group experienced a so-called needle-

[1] VOR refers to a very high frequency radio beacon that is used for navigation.

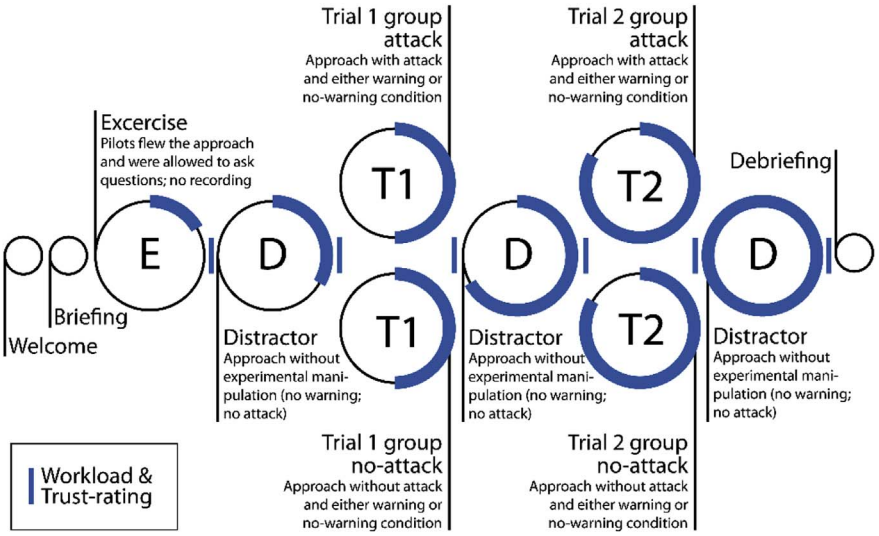

Fig. 4. Experimental procedure.

swing of the MDF VOR 8 nm before MDF. Although a needle-swing (the VOR needle's rapid 180° direction change) indicating flight over a VOR tells the pilots that they have just passed MDF, this was not so at D8MDF. Since the pilots have to turn left 3.8 nm before MDF, they would not expect any needle-swing on the MDF-tuned VOR during the whole approach. We chose to manipulate that specific system as we wanted to pose a problem that could be rather easily handled by the pilots. The pilots could fly below the cloud layer to acquire visual reference and use other VOR stations. Even if the pilots were not able to solve the problem, they could still fly the aircraft. Furthermore, we wanted to design the manipulation to imitate realistic aircraft behavior (a needle swing is the actual behavior of an aircraft flying over a VOR station). Our intention was to create just enough difficulty to defeat about half the pilots, thus avoiding floor and ceiling effects, and also producing meaningful sample sizes in the successful- and unsuccessful-pilot groups.

2.4.5. Warning condition

Under the warning condition, the air traffic controller gave a warning 11 nm before MDF. The air traffic controller contacted the aircraft and communicated that “it is suspected that your aircraft could be under a cyber-attack that might have compromised your systems” and cleared the aircraft for landing on runway 26L. The air traffic controller further stated that no more radio communication was allowed for *security reasons* to preclude further information requests from the pilots. In the attack group, the warning condition helped the pilots to understand the situation, representing a *hit*, whereas in the no-attack group, pilots experienced a *false alarm*. Under no-warning conditions the attack group experienced a *miss*, the no-attack group a *correct rejection*, which can be seen as a baseline.

2.4.6. Flight-information package

The paper-based flight-information package contained a description and instructions for use of the simulator (instrument equipment list with detailed explanation), initial navigation setting, an approach map as well as pitch-power settings, and current weather information. To ensure that the simulator set-up was the same for every pilot and every trial, pilots had to work through a simulator-preparation checklist included in the package before the beginning of every trial.

2.5. Procedure

The study was performed within a period of 43 days during which the 22 participants were tested. The experiment took between 120 min and 150 min.

Participants filled out the demographic questionnaire after they were welcomed and had given their written consent to participation in the experiment. After that, participants received the flight-information package and were briefed on the simulator and the upcoming task (see Fig. 4 for an overview of the procedure). Pilots were told that the experiment concerned a general ergonomic issue involving cockpit design. Participants next had a chance to get familiar with the simulator and the approach via an exercise during which they were allowed to ask questions. The eye-tracking system was configured and calibrated afterward.

Participants flew the first scenario after calibration. Since there was no experimental manipulation, we called it a *distractor*. The pilots subsequently flew the experimental scenario and either did or did not receive a warning during *Trial 1*. Following another distractor scenario, participants flew *Trial 2*. The experiment ended with another distractor scenario. After each scenario, pilots completed the NASA *r*TLX and trust questionnaire. After completing all of the scenarios, the pilots filled out the qualitative questionnaire and were debriefed by the investigators. Participants then received travel compensation and had a chance to participate in a lottery.

2.6. Data quality and processing

2.6.1. Eye-tracking

The pilots' gaze data was checked to ensure that it met the data-availability quality criteria defined in ISO/TS 15007–2:2014–09. At least 85% of the frames have to have valid pupil detection to achieve *good* quality; 95% are needed for *excellent* quality. Pupil detection was excellent (M = 97.6%) on average with the lowest value being 93.3%. The eye-tracker was recalibrated before computing gaze statistics in case the head-based eye-tracker had been dislocated on any of the participants' heads. Gaze statistics were calculated using D-Lab software (version 3.10.7757).

2.6.2. Flight data

Simulator flight data was collected using MATLAB Simulink 2015b. The data sample rate was set to 100 Hz.

2.6.3. Data exclusion

We had to exclude one participant from all analyses as he did not meet the test-person requirements (see 2.2). This left 21 participants for the performance measurement. We also had to exclude data from further subjects due to technical problems, which left 20 data sets for the

analysis of workload and trust and 19 data sets for the eye-tracking analyses.

3. Results

3.1. Statistical analysis

Statistical analyses were performed on a significance level of $\alpha = .05$. We used two-way mixed-model analysis of variance (ANOVA) to test for differences between (factor *attack*) and within (factor *warning*) the groups. *P*-values are reported two-sided unless stated otherwise; partial eta-squares are used as measures of effect size with $\eta^2_{partial} > .14$ regarded as a large effect. As several cells in the contingency tables contained observed frequencies of less than 5, we did not calculate chi-square statistics, but rather applied Fisher's exact probability test and calculated *Odd's ratio* and the *Relative Risk*. Error bars in the figures denote to ± 1 standard deviation.

3.2. Workload

The two-way mixed ANOVA on subjectively perceived workload (see Fig. 5) shows no main effect of the warning, $F(1,18) = 1.03$, $p = .32$, but a significantly large effect from the attack, $F(1,18) = 4.76$, $p = .04$, $\eta^2_{partial} = .209$, so that the workload is perceived to be heavier in the attack-group than it is in the no-attack group. Further, the large interaction effect was found to be significant, $F(1,18) = 13.00$, $p = .002$, $\eta^2_{partial} = .419$.

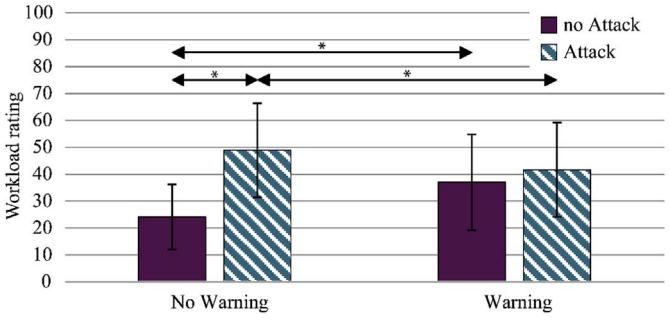

Fig. 5. Subjective workload rating of the two groups (no-attack and attack) under the two experimental conditions (no-warning and warning).

We performed Bonferroni-corrected pairwise comparisons showing significant differences under the no-warning condition between the no-attack ($M = 24.10$, $SD = 3.65$) and attack group ($M = 48.90$, $SD = 5.85$), $p = .001$, but no differences under the warning condition, $p = .561$. Perceived workload was significantly heavier in the no-attack group under the warning condition ($M = 36.95$, $SD = 5.35$) than under the no-warning condition, $p = .003$. The attack group's members perceived heavier workload under the no-warning condition than under the warning condition ($M = 41.65$, $SD = 5.85$), $p = .049$ (one-tailed).

3.3. Trust

The trust rating (see Fig. 6) shows no significant difference whether there is a warning or not, $F(1,18) = .00$, $p = .996$. The attack affects the participants' trust, $F(1,18) = 8.68$, $p = .009$, $\eta^2_{partial} = .325$, so that members of the no-attack group exhibit stronger overall trust in the aircraft's systems than do members of the attack group.

The interaction effect between the warning and attack factor was found to not be significant, $F(1,18) = 2.66$, $p = .121$. The no-attack group's trust rating ($M = 5.49$; $SD = .22$) was significantly greater than that of the attack group ($M = 3.95$; $SD = .34$), $p < .001$, under the no-warning condition. No other pairwise comparisons were found to be significant.

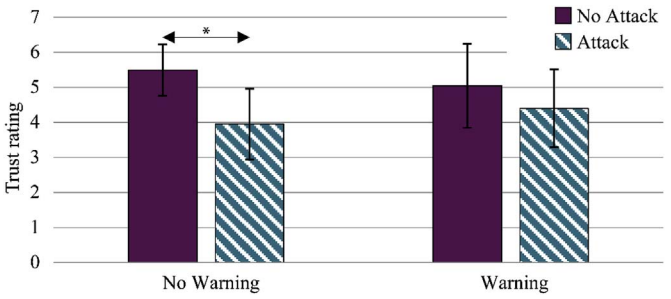

Fig. 6. System trust of the two groups (no-attack and attack) under the two experimental conditions (no-warning and warning).

3.4. Eye-tracking

The analysis shows an attention distribution as depicted in Fig. 7 for the three defined AOI-sets: *Aviate*, *navigate*, and *system management*. The two-factorial mixed ANOVA on attention ratio on the *aviate* set shows a large, statistically significant main effect of attack, $F(1,17) = 7.07$, $p = .017$, $\eta^2_{partial} = .294$ giving the no-attack group a higher attention ratio on the *aviate* set than that of the attack group. The warning's effect was found to be non-significant, $F(1,17) = 1.65$, $p = .216$. A large and statistically significant interaction effect was found, $F(1,17) = 4.57$, $p = .047$, $\eta^2_{partial} = .212$. Pairwise comparison showed a significant difference between the two groups where the no-attack group had a higher attention ratio ($M = 59.67$, $SD = 6.60$) than the attack group ($M = 35.21$; $SD = 4.29$) under the no-warning condition, $p = .008$. The no-attack group showed a significantly higher attention ratio under the no-warning condition ($M = 59.67$, $SD = 6.60$) than under the warning condition ($M = 39.39$, $SD = 4.53$), $p = .024$.

A two-factorial mixed ANOVA on attention ratio on the *navigate* set shows a large, statistically significant main effect of the attack, $F(1,17) = 16.21$, $p = .001$, $\eta^2_{partial} = .488$, and a non-significant effect of the warning, $F(1,17) = 1.45$, $p = .245$. Further, the interaction was non-significant, $F(1,17) = .11$, $p = .743$. Participants in the attack group had a significantly higher attention ratio on the *navigate* set than did the no-attack group. Pairwise comparisons showed significant

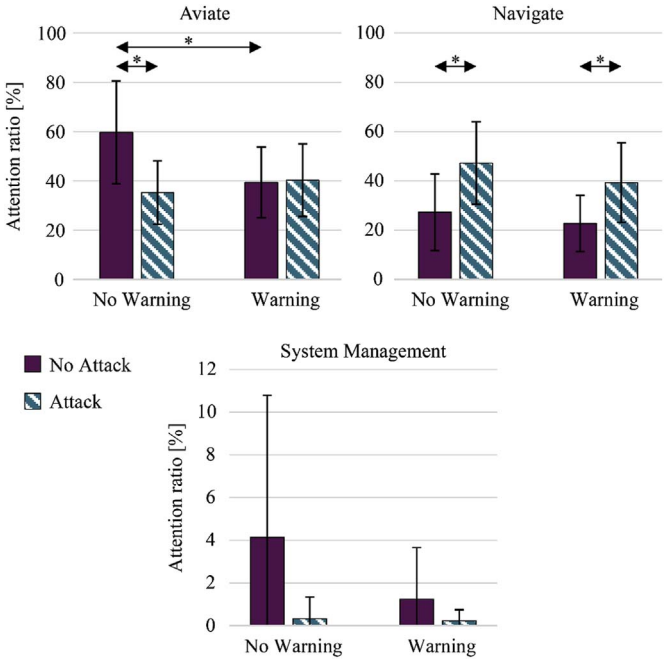

Fig. 7. Attention ratio of the two groups (no-attack and attack) on the *aviate, navigate,* and *system management* sets under the two experimental conditions (no-warning and warning). Note the lower and upper graphs' different scales.

differences under the no-warning condition between the attack group (M = 47.26, SD = 5.58) and the no-attack group (M = 27.26, SD = 4.91), p = .019, as well as under the warning condition. Here, the attack group showed a significantly higher attention ratio (M = 39.25, SD = 5.41) on the *navigate* set than did the no-attack group (M = 22.73, SD = 3.61), p = .015. A third mixed ANOVA on the attention ratio on the *system management* set was conducted showing a large significant effect of the attack, $F(1,17) = 4.79$, p = .043, $\eta^2_{partial}$ = .220, and a non-significant effect of the warning, $F(1,17) = 1.35$, p = .261. The interaction effect was also found to be non-significant, $F(1,17) = 1.17$, p = .294. As these data are not normally distributed, we did not conduct any further pairwise comparisons. Since ANOVAs have shown to be robust against violations of normal distribution (Schmider et al., 2010), we do not anticipate any problems when interpreting the reported effects.

3.5. Performance

Figure 8 shows the participants' flight paths and the 3.8 nm circle around the MDF VOR. Turns initiated before that circle were regarded as unsuccessful and turns after the 3.8 nautical mile circle as successful approach decisions.

The performance classification leads to the contingency table shown in Table 2. Fisher's exact probability test shows that the group under attack fails significantly more often than the group not under attack, p (two-tailed) = .04. The *Odd's ratio* shows a medium effect with a value of 5.2; the *R*elative *R*isk is 3.3.

As the observed frequency for both warning conditions is the same, we conclude that the warning has no main effect on pilots' decision quality. Fisher's exact probability test for comparing warning effects within one group was found to be not significant.

4. Discussion

4.1. General points

The results showed a significant effect of the attack in all analyses, whereas the effect of the warning was most obvious in a strong interaction effect. The effect of the attack can primarily be explained by the additional effort and intense hypotheses testing pilots have to engage in. The warning, in contrast, was shown to be a strong moderator in the form of large interaction effects. This can also be explained in terms of consistency: whereas information from the air traffic controller and the actual aircraft state is consistent under *hit* and *correct rejection* conditions, the information is inconsistent for *false alarms* and *misses*.

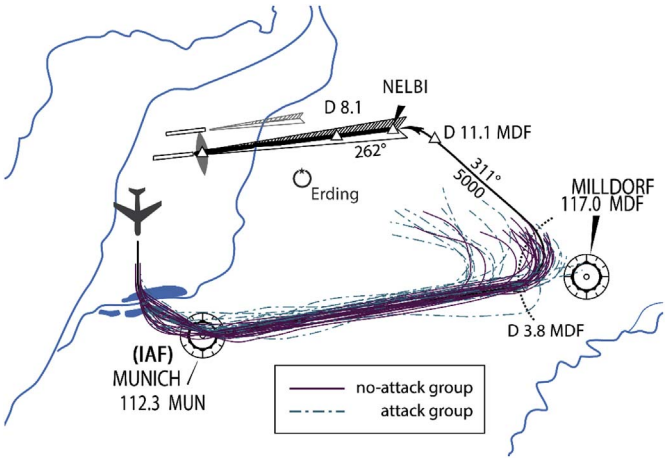

Fig. 8. The 21 participants' flight paths including the 3.8 nm circle around the MDF VOR at which pilots had to turn in this scenario. Dashed flight paths signify the attack group; solid flight paths denote to the no-attack group. Reduced map excerpt is for the purpose of illustration only. NOT FOR NAVIGATION.

Table 2
Contingency table for successful and unsuccessful approach decisions.

	Warning	No-warning	Sum
Attack	7 (3)	4 (6)	11 (9)
No-attack	8 (3)	11 (0)	19 (3)
Sum	15 (6)	15 (6)	30 (12)

4.2. Workload (RQ 1 and RQ 2)

The results pertaining to the workload rating show that the warning has no main effect, but that the attack does. Furthermore, an interaction effect is present. This leads to the conclusion that the attack evokes cognitive activation as described in the deception detection model (Grazioli, 2004), which increases the workload. This result is consistent with previous research by Hirshfield et al. (2015), who also reported increased workload when users are under cyber-attack.

RQ 1: The presence of a cyber-attack increases pilots' workload.

This is especially true for *misses*. *False alarms* and *hits* both entail heavier workload than *correct rejections*; *misses*, however, show the highest rating. We argue that an appropriate warning can lighten the workload, but want to point out that a *false alarm* is about as demanding as a *hit*. This relationship can also be explained by the *cue-clarity* model (Orasanu et al., 1993), which shows that cues associated with ambiguous problems (such as the attack in our case) require more cognitive work than do those associated with unambiguous problems (like low fuel). *Correct rejection*, which can also be seen as a baseline, features neither an attack nor a warning so that pilots only have to complete the flight task without any additional hypotheses generation or decision-making. *False alarms* cause cognitive activation in the pilots and are likely to increase the workload until he or she has checked all instruments and ruled out an attack. *Hits* and *misses* require the pilots to make decisions thereby increasing the workload. Inconsistency further exacerbates this effect for a *miss*.

The VOR needle-swing and the pilot's lack of information about the underlying situation rendered the aircraft's state ambiguous. A correct warning helped the pilots, since it gave them an explanation for the ambiguous state of the aircraft. In contrast, a *false alarm* created ambiguity, because the pilots did not know the cause of the alarm. We can conclude that consistent warnings can lighten pilots' workload but inconsistent ones increase it.

RQ 2: An attack warning lightens pilots' workload for a *hit* and increase it for a *false alarm*.

Based on the results here, one could argue that *false alarms* are better than *misses*; however, we want to point out that the pilots experienced only one *false alarm* or *miss* in our study. One has to design the alarm system to carefully avoid evoking cry-wolf effects (e.g., Bliss et al., 1995) caused by excessive sensitivity thereby weakening trust in the system finally resulting in its disuse.

4.3. Trust (RQ 3 and RQ 4)

The attack group exhibited less trust than did the no-attack group, which we attribute to the system's perceived unreliability. An analysis from Bliss (2003) showed that about 40% of alarm-related events in the *Aviation Safety Reporting System* database are either *false alarms* or *misses*. That lower perceived technical reliability weakens trust in the system is already known from several other domains (see section 1.2) and culminates in pilots deactivating the alarm systems (Sorkin, 1988).

RQ 3: The presence of a cyber-attack weakens pilots' trust in the system.

Trust is weakest when pilots are confronted with a *miss*, whereas a *false alarm* does not seem to influence the trust rating on its first occurrence. It seems that a *miss* has a larger negative effect on pilot's trust than a *false alarm* has. We are nevertheless of the opinion that higher occurrence rates of *false alarms* would considerably reduce the trust rating —a manifestation of the cry-wolf effect (see also Manzey et al. (2014) and Wickens et al. (2009) for further discussion). Furthermore, the missing cry-wolf effect in our setting might be due to the design of the warning which was enunciated with a certain degree of uncertainty about the presence of a threat (ATC: "… aircraft could be under attack …").

RQ 4: A warning of an attack did not attenuate the attack's effect on pilots' trust.

Detailed information about the confidence of the alarm system itself, as discussed by Antifakos et al. (2005), might positively influence pilots' trust. The trust ratings obtained seem to be inversely related to the workload ratings. It seems that the heavier the perceived workload, the weaker the trust in the system is and vice versa. This relation might be explained by pilots' flying experience: In daily operation, pilots perceive highest workload when there is a technical problem which is attributable to reduced system reliability and leads to reduced trust.

4.4. Eye-tracking (RQ 5 and RQ 6)

We observed an elevated attention ratio of about 60% in case of *correct rejection* toward *aviate* as well as about 25% toward *navigate*. For *false alarm*, the warning seems to shift attention from *aviate* to other areas. Although this difference is not statistically significant, it shows that pilots are checking further instruments to rule out an attack. The air traffic controller's warning seems not to affect the *navigate* set, because the former gives no further information about either the source of the problem or the affected systems. The attack group evinces a greater effect than the no-attack group in that the attention ratio is significantly larger in the *navigate* set. This effect is attributable to the attack primarily affecting the navigational task and not the flying or stabilizing of the airplane (*aviation* set) per se. For the *management* set, we expected to see a very different picture where the presence of a warning or an attack would cause the attention ratio to be higher than in the *correct rejection* (baseline). However, the contrary relation seems to prevail in that the warning and the presence of an attack absorb so much attention that there is none left for *management* activities. Such attention tunneling might contribute to greater risk because necessary monitoring of the basic flying instruments might no longer be given (Dehais et al., 2015; Tessier and Dehais, 2012).

RQ 5: A genuine cyber-attack influences pilots' attention ratio across all analyzed sets of areas because attention shifts toward problem-solving and monitoring of basic instruments might be neglected.
RQ 6: A cyber-attack *false alarm* leads to a decrease in monitoring basic flying instruments.

Ongoing analyses focus on the relation between cue consistency and gaze behavior, trust, and workload estimates as well as on their interdependencies. Based on the studies of Hergeth et al. (2016) and Oakley et al. (2003), we hypothesize a negative relationship between attention ratio and trust in the system; system reliability influences the latter.

4.5. Performance (RQ 7 and RQ 8)

The results showed that the success rate of the attack group was significantly lower than that of the no-attack group—regardless of the warning. We nevertheless see that without a warning, more than half of the pilots in the attack group thought they had overshot the VOR and immediately turned left.

RQ 7: A cyber-attack leads to more incorrect decisions.

Decision-making quality can be treated as a performance measure. Pilots' infrequent need to make decisions in situations lacking procedural guidance also seems to explain the rather low success rates. The pilots' scant opportunity to train and strengthen their decision-making skills for coping with uncertainty during unforeseen events further depresses the success rate. This is also in line with the results of Gontar et al. (2015), who found that pilots are mainly trying to apply analytical decision-making strategies (as they are taught to) rather than making recognition-primed decisions based on their experience. We argue that while under cyber-attack, recognition-primed decisions should be expected to be especially rare because pilots are not currently trained in how to respond to such unforeseen events nor do they have a chance to accumulate experience in handling them. This combined with the lack of procedures seems to make it even harder for the pilots to react appropriately.

RQ 8: Consistent warnings tend to help the pilots but inconsistent ones further decrease their performance.

Warnings show no significant main effect on pilots' performance. No significant difference in the influence of warnings was found within groups. Nevertheless, we believe that consistent warnings might help the pilots whereas inconsistent ones further confuse them. A *false alarm* confused three participants in the no-attack group causing them to decide incorrectly about when to turn. This might indicate that navigational tasks are already severely challenging some pilots because they are flying their aircraft normally with flight-director and autopilot engaged (Haslbeck and Hoermann, 2016). Raw data flying—flying according to the needles—seems to challenge some of the pilots even without any further tasks, but especially when confronted with a secondary task such as verifying their instruments in an ambiguous, unexpected situation.

4.6. Qualitative data

When we asked our participants what "would have helped you in this situation?" in a qualitative questionnaire, they answered that deeper knowledge of the system and understanding of which systems are generally vulnerable to cyber-attacks would have helped them. They further stated that more practice would have supported them during the manual flying task. Other participants stated that an independent, non-electrical back-up system would be of great support along with a virus scanner equipped to detect potential intrusion. The qualitative data obtained prompts us to conclude that pilots are not only unable to hypothesize that they are under attack, but that their knowledge about how to respond is also inadequate. Answers to the question "What information would you like to have had in the air traffic controller message?" also reflects this. Here, pilots stated that they wanted to have more information about the cyber-attack itself, which aircraft components are affected, and what consequences this might have.

5. Limitations of the study

One of the study's limitations is its rather small sample size of 22 participants. Another is the very heterogeneous distribution of experience, as we did not control for experience effects and randomly assigned participants to either of the groups. We were nonetheless able to ensure that the sample's broad heterogeneity covered diverse aspects of operator response, which we think is important in an exploratory study.

We are, however, most concerned about the overrepresentation of *false alarms* and *misses*, because a well-functioning system would exhibit a smaller probability of their occurrence. Although we do not know how the pilots' behavior would change after experiencing several *false alarms*, we do anticipate cry-wolf effects here. As we did not manipulate the alarms' ambiguity and degree of uncertainty, we cannot conclude how these factors influence the pilots' behavior. Further research addressing these specific aspects is needed. Moreover, future researchers should manipulate the complexity and severity of the cyber-attacks. Such variations constitute the number of manipulated systems or flight instruments involved to determine whether our observed effects are invariant across different types of attacks. Although we implemented a scenario that was rather easy to deal with, we found a considerable number of pronounced effects. Specifically, we assume very high drop-out rates and fewer conclusions to be drawn when presenting very complex scenarios or ones that are unmanageable for the pilots.

6. Conclusion and recommendations

Our results show that cyber-attacks influence pilots' workload, trust in the system, visual information acquisition behavior, and performance. We were able to show that a warning about an impending cyber-attack can moderate several of those effects but cannot completely obviate them. Thus the cyber-attack and the warning have to be taken into account when one wants to attenuate the aforementioned effects.

6.1. Blunting the impact of the attack effect

Of course, technical and organizational efforts should be expended to establish every possible barrier for reducing the probability of a successful cyber-attack. As we argue, even the most sophisticated barriers cannot completely exclude the possibility of successful breaches, which the pilots then have to handle. From a human-factors perspective, one very important outcome of this study was that pilots did not take any cyber-attacks into account when they tried to solve the problem they faced. From a decision-making viewpoint, the mere variety of potential attacks seems to render pattern recognition nearly impossible. Since different underlying manipulations can even introduce the same errors from the operator's perspective, it even seems detrimental if pilots apply their experience, as the cues that they perceive might be invalid in a given context and can change from case to case. Taking the qualitative data into account, we recommend incorporating the following approaches to diminish the effect of the attack itself (see also International Federation of Air Line Pilots' Associations (2013) for further training approaches):

1) **Cyber-attack awareness training:** Pilots need training that raises their awareness of potential threats and of how an aircraft can be infected with malicious software. This should also include aspects of everyday behavior such as connecting a mobile phone to the electronic flight bag's USB port for charging, which can be sufficient to infect the aircraft. Although we concentrated mainly on the pilots in this article, we believe that such training is also very important for other staff members such as the cabin or maintenance personal to establish “collective awareness of cyber threats” (International Civil Aviation Organization, 2016, p. 2). One approach to such training might be to train pilots using simulation games in the classroom, or using unforeseen events in the simulator. Since we generally recommend incorporating unforeseen events into simulator training, doing so in this context might also be helpful.
2) **General decision skill training**: Even if pilots are aware of the potential threat, they need training that helps them to resolve the situation. Although not even a fraction of all possible attack scenarios can be trained, the underlying skills should be. That is, pilots should be trained in handling unforeseen and unexpected events. Such training could also be incorporated into current training syllabi as we are aware of cost pressure and acknowledge that several hours of extra training for each pilot are infeasible. Another approach involves using tools such as *ShadowBox* (Klein and Borders, 2016). These authors acknowledge that subject-matter experts “are good at the skill they are teaching but may not be good at teaching that skill” (Klein and Borders, 2016, p. 268) and also that the requisite subject-matter experts are very expensive. When using this tool, trainees are presented with an unusual scenario in form of narratives, video, or audio material (Mosier et al., 2018). At specific decision points, trainees are asked to prioritize different decision aspects such as actions to take, information selection, and cue interpretation. They subsequently give the rationale for their choice. The trainees' choices and rationales are then compared to those of subject-matter experts (Klein and Borders, 2016). Such an approach could facilitate the teaching of decision-making skills for unforeseen events in a very general and thus useful way while incurring only minimal cost.

6.2. Blunting the impact of the warning effect

Our research showed how warnings profoundly influenced different operator characteristics. In a recent analysis of alerting systems' role in aviation accidents, Mumaw (2017) found that pilots quite often failed to detect or even understand incoming visual or auditory alerts (see also Bliss, 2003). Poorly designed alarms are particularly worrisome in that they can greatly increase stress or distract pilots (Peryer et al., 2005; Doll et al., 1983) thereby failing to establish sufficient awareness—an effect that is especially prominent during flight phases that put severe demands on the pilots (Durantin et al., 2017). In contrast, correct and trustworthy alarms can reduce not only pilots' cognitive load but also establish an appropriate picture of the situation and hence enhance their performance. In line with the results of the qualitative data, we recommend implementing warning systems based on systems comparable to virus scanners or firewalls.

3) **Cyber-attack warnings:** Cyber-attack warning systems comparable to virus-scanners or firewalls designed to protect computers should be installed to inform pilots when parts of the system have been infiltrated. To diminish the cry-wolf effects, such warning systems should be designed as likelihood alarms (see Sorkin, 1988). That is, a well-designed alarm not only communicates which system might be affected, but also how certain the threat is. Ideally, the warning system would also suggest different options for handling the impending threat including a very brief rationale for each option. Such alarms could enhance the salience, importance, and meaning of the relevant parameters, thus supporting pilots in providing clear cues fostering good decision-making. We are very aware that such warning systems can substitute for neither pilot decision-making nor awareness training. The warning system can nonetheless represent an altogether valuable support tool for pilots.

With the rising number of threats, there will also be an increased number of attacks against civil aircraft. As long as we have pilots flying our aircraft, they will be the last line of defense. We therefore have to support them with procedures and training in every possible way so that they are aware of, can correctly detect, and appropriately handle impending cyber-attacks. Such approaches could be implemented in form of a security management system analogous to well-established safety management systems in every airline. The research reported here takes a first step into the human-factors related research necessary for this endeavor.

Acknowledgements

This paper originates from an interdisciplinary project of HH and MR together with Anna Wegleiter and Claudius Wilhelm, whose support during the experiment and the analyses we sincerely acknowledge. We thank Immanuel Barshi for fruitful discussions about and comments on earlier versions of the manuscript. Special thanks goes to Christoph Krause and Chong Wang from the Chair of Flight System Dynamics for their technical support in the flight simulator and to Josef Niederl for supporting us in the development of the flight scenario.

References

American Institute of Aeronautics and Astronautics, 2013. The Connectivity Challenge: Protecting Critical Assets in a Networked World: a Framework for Aviation Cybersecurity.

Antifakos, S., Kern, N., Schiele, B., Schwaninger, A., 2005. Towards improving trust in context-aware systems by displaying system confidence. In: Tscheligi, M., Bernhaupt, R., Mihalic, K. (Eds.), Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services. ACM, New York, NY, pp. 9–14. https://doi.org/10.1145/1085777.1085780.

Bahner, J.E., Hüper, A.-D., Manzey, D., 2008. Misuse of automated decision aids: complacency, automation bias and the impact of training experience. Int. J. Hum. Comput. Stud. 66 (9), 688–699. https://doi.org/10.1016/j.ijhcs.2008.06.001.

Bergström, J., Dahlström, N., van Winsen, R., Lützhöft, M., Dekker, S., Nyce, J., 2014. Rule- and role-retreat: an empirical study of procedures and resilience. J. Marit. Res. 6 (1).

Biesecker, C., 2017, August 11. DHS led team demonstrates that commercial aircraft can Be remotely hacked. Defense Dail Retrieved from: http://www.defensedaily.com/dhs-led-team-demonstrates-commercial-aircraft-can-remotely-hacked.

Bliss, J.P., 2003. Investigation of alarm-related accidents and incidents in aviation. Int. J. Aviat. Psychol. 13 (3), 249–268. https://doi.org/10.1207/S15327108IJAP1303_04.

Bliss, J.P., Dunn, M.C., 2000. Behavioural implications of alarm mistrust as a function of task workload. Ergonomics 43 (9), 1283–1300. https://doi.org/10.1080/001401300421743.

Bliss, J., Dunn, M., Fuller, B.S., 1995. Reversal of the cry-wolf effect: an investigation of two methods to increase alarm response rates. Percept. Mot. Skills 80 (3c), 1231–1242. https://doi.org/10.2466/pms.1995.80.3c.1231.

Bobko, P., Barelka, A.J., Hirshfield, L.M., 2014. The construct of state-level suspicion. Hum. Factors 56 (3), 489–508. https://doi.org/10.1177/0018720813497052.

Boyce, M.W., Duma, K.M., Hettinger, L.J., Malone, T.B., Wilson, D.P., Lockett-Reynolds, J., 2011. Human performance in cybersecurity: a research agenda. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 55 (1), 1115–1119. https://doi.org/10.1177/1071181311551233.

Breznitz, S., 1984. Cry Wolf: the Psychology of False Alarms. Lawrence Erlbaum Associates, Hillsdale, N.J.

Byers, J.C., Bittner, A.C., Hill, S.G., 1989. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In: Mital, A. (Ed.), Advances in Industrial Ergonomics and Safety I. Taylor & Francis, New York, NY, pp. 481–485.

Caldwell, J.A., 2012. Crew schedules, sleep deprivation, and aviation performance. Curr. Dir. Psychol. Sci. 21 (2), 85–89. https://doi.org/10.1177/0963721411435842.

Casner, S.M., Geven, R.W., Williams, K.T., 2013. The effectiveness of airline pilot training for Abnormal events. Hum. Factors J. Hum. Factors Ergon. Soc. 55 (3), 477–485. https://doi.org/10.1177/0018720812466893.

Dahlström, N., Dekker, S., van Winsen, R., Nyce, J., 2009. Fidelity and validity of simulator training. Theor. Issues Ergon. Sci. 10 (4), 305–314. https://doi.org/10.1080/14639220802368864.

Dehais, F., Tessier, C., Chaudron, L., 2003. GHOST: experimenting conflicts countermeasures in the pilot's activity. In: Proceedings of the 18th. Morgan Kaufmann Publishers, San Francisco, CA, pp. 163–168.

Dehais, F., Causse, M., Pastor, J., 2008. Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring. In: Federal Aviation Administration (Ed.), Proceedings of the Third International Conference on Research in Air Transportation, pp. 303–309 Fairfax, VA.

Dehais, F., Tessier, C., Christophe, L., Reuzeau, F., 2010. The perseveration syndrome in the Pilot's activity: guidelines and cognitive countermeasures. In: Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J.M., Mattern, F., Mitchell, J.C., Winckler, M. (Eds.), Lecture Notes in Computer Science. Human Error, Safety and Systems Development. vol. 5962. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 68–80. https://doi.org/10.1007/978-3-642-11750-3_6.

Dehais, F., Peysakhovich, V., Scannella, S., Fongue, J., Gateau, T., 2015. Automation surprise in aviation: real-time solutions. In: Kim, J. (Ed.), Proceedings of the 33rd Annual CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, pp. 2525–2534.

Dismukes, K., Berman, B.A., Loukopoulos, L.D., 2007. The limits of expertise: rethinking pilot error and the causes of airline accidents. In: Key Dismukes, R., Berman, Benjamin A., Loukopoulos, Loukia D. (Eds.), Ashgate Studies in Human Factors for Flight Operations. Ashgate, Aldershot.

Doll, T.J., Folds, D.J., Leiker, L.A., 1983. Auditory Information Systems in Military Aircraft: Current Configurations Versus State of the Art (Report USAFSAM-TR-84–15). Brooks Air Force Base, TX.

Durantin, G., Dehais, F., Gonthier, N., Terzibas, C., Callan, D.E., 2017. Neural signature of inattentional deafness. Hum. Brain Mapp. 38 (11), 5440–5455. https://doi.org/10.1002/hbm.23735.

Dutt, V., Ahn, Y.-S., Gonzalez, C., 2013. Cyber situation awareness. Hum. Factors 55 (3), 605–618. https://doi.org/10.1177/0018720812464045.

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G., Beck, H.P., 2003. The role of trust in automation reliance. Int. J. Hum. Comput. Stud. 58 (6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7.

Elias, B., 2015. Protecting Civil Aviation from Cyberattacks.

European Aviation Safety Agency, 2016. The European Plan for Aviation Safety. Cologne.

Federal Aviation Administration, 2013. Special conditions: Boeing model 777–200, –300, and –300ER series airplanes: aircraft electronic system security protection from unauthorized internal access. Fed. Regist. 78 (222), 68985–68986.

Federal Aviation Administration, 2014. Special conditions: Airbus model A350–900 airplanes: isolation or protection of the aircraft electronic system security from unauthorized internal access. Fed. Regist. 79 (143), 43239–43240.

Firesmith, Donald G., 2003. Common Concepts Underlying Safety, Security, and Survivability Engineering.

Fox, S.J., 2016. Flying challenges for the future: aviation preparedness – in the face of cyber-terrorism. J. Transport. Stat. 9 (3–4), 191–218. https://doi.org/10.1007/s12198-016-0174-1.

Gold, C., Körber, M., Hohenberger, C., Lechner, D., Bengler, K., 2015. Trust in automation – before and after the experience of take-over scenarios in a highly automated vehicle. Proc. Manuf. 3, 3025–3032. https://doi.org/10.1016/j.promfg.2015.07.847.

Gontar, P., Hoermann, H.-J., 2015. Interrater reliability at the top end: measures of pilots' nontechnical performance. Int. J. Aviat. Psychol. 25 (3–4), 171–190. https://doi.org/10.1080/10508414.2015.1162636.

Gontar, P., Mulligan, J.B., 2015. A metric to quantify shared visual attention in two-person teams. In: Pfeiffer, T., Essig, K. (Eds.), Proceedings of the 2nd International Workshop on Solutions for Automatic Gaze Data Analysis 2015 (SAGA 2015), pp. 37–38 Bielefeld.

Gontar, P., Mulligan, J.B., 2016. Cross recurrence analysis as a measure of pilots' coordination strategy. In: Droog, A., Schwarz, M., Schmidt, R. (Eds.), Proceedings of the 32nd Conference of the European Association for Aviation Psychology, pp. 524–544 Groningen, NL.

Gontar, P., Porstner, V., Hoermann, H.-J., Bengler, K., 2015. Pilots' decision-making under high workload: recognition-primed or not – an engineering point of view. In: Lindgaard, G., Moore, D. (Eds.), Proceedings of the 19th Triennial Congress of the International Ergonomics Association, Melbourne.

Gontar, P., Fischer, U., Bengler, K., 2017a. Methods to evaluate pilots' cockpit communication: cross-recurrence analyses vs. Speech act-based analyses. J. Cognit. Eng. Decis. Making 78 (1). https://doi.org/10.1177/1555343417715161.

Gontar, P., Schneider, S.A.E., Schmidt-Moll, C., Bollin, C., Bengler, K., 2017b. Hate to interrupt you, but… analyzing turn-arounds from a cockpit perspective. Cognit. Technol. Work 11 (2), 130. https://doi.org/10.1007/s10111-017-0440-4.

Grazioli, S., 2004. Where did they go Wrong? An analysis of the failure of knowledgeable internet consumers to detect deception over the internet. Group Decis. Negot. 13 (2), 149–172. https://doi.org/10.1023/B: GRUP.0000021839.04093.5d.

Haass, J., Sampigethaya, R., Capezzuto, V., 2016. Aviation and cybersecurity: opportunities for applied research. TR News 304, 39–43.

Haslbeck, A., Bengler, K., 2016. Pilots' gaze strategies and manual control performance using occlusion as a measurement technique during a simulated manual flight task. Cognit. Technol. Work 18 (3), 529–540. https://doi.org/10.1007/s10111-016-0382-2.

Haslbeck, A., Hoermann, H.-J., 2016. Flying the needles: flight deck automation erodes fine-motor flying skills among airline pilots. Hum. Factors 58 (4), 533–545. https://doi.org/10.1177/0018720816640394.

Haslbeck, A., Zhang, B., 2017. I spy with my little eye: analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Appl. Ergon. 63, 62–71. https://doi.org/10.1016/j.apergo.2017.03.015.

Haslbeck, A., Schubert, E., Gontar, P., Bengler, K., 2012. The relationship between pilots' manual flying skills and their visual behavior: a flight simulator study using eye tracking. In: Laundry, S., Salvendy, G., Karwowski, W. (Eds.), Advances in Human Factors and Ergonomics, Advances in Human Aspects of Aviation. CRC Press, Boca Raton, pp. 561–568.

Heereman, J., Walla, P., 2011. Stress, uncertainty and decision confidence. Appl. Psychophysiol. Biofeedback 36 (4), 273–279. https://doi.org/10.1007/s10484-011-9167-9.

Heiges, M., Bever, R., Carnahan, K., 2015. How to safely flight test a UAV subject to cyber-attacks. In: Presentation at SCI-269 Symposium. NATO Science and Technology Organization (STO).

Hergeth, S., Lorenz, L., Vilimek, R., Krems, J.F., 2016. Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. Hum. Factors 58 (3), 509–519. https://doi.org/10.1177/0018720815625744.

Hirshfield, L., Bobko, P., Barelka, A.J., Costa, M.R., Funke, G.J., Mancuso, V.F., Knott, B.A., 2015. The role of human operators' suspicion in the detection of cyber attacks. Int. J. Cyber Warf. Terror. (IJCWT) 5 (3), 28–44. https://doi.org/10.4018/IJCWT.2015070103.

Horowitz, B.M., Lucero, D.S., 2016. System-aware cyber security: a systems engineering approach for enhancing cyber security. INSIGHT 19 (2), 39–42. https://doi.org/10.1002/inst.12087.

Iasiello, E., 2014. Aviation and Cyberspace: Convergence of Domains, Convergence of Threats.

International Civil Aviation Organization, 2012. Cyber Security for Civil Aviation. Montreal.

International Civil Aviation Organization, 2016. Coordinating Cybersecurity Work. Montreal.

International Federation of Air Line Pilots' Associations, 2013. Cyber Threats: Who Controls Your Aircraft?.
Jian, J.-Y., Bisantz, A.M., Drury, C.G., 2000. Foundations for an empirically determined scale of trust in automated systems. Int. J. Cognit. Ergon. 4 (1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04.
Klein, G., 1999. Sources of Power: How People Make Decisions, second ed. MIT press, Cambridge, MA, London.
Klein, G., Borders, J., 2016. The ShadowBox approach to cognitive skills training. J. Cognit. Eng. Decis. Making 10 (3), 268–280. https://doi.org/10.1177/1555343416636515.
Langton, J.T., Baker, A., 2013. Information visualization metrics and methods for cyber security evaluation. In: IEEE International Conference. Intelligence and Security Informatics (ISI), pp. 292–294. https://doi.org/10.1109/ISI.2013.6578846.
Lee, J., Moray, N., 1992. Trust, control strategies and allocation of function in human-machine systems. Ergonomics 35 (10), 1243–1270. https://doi.org/10.1080/00140139208967392.
Lee, J.D., See, K.A., 2004. Trust in automation: designing for appropriate reliance. Hum. Factors J. Hum. Factors Ergon. Soc. 46 (1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392.
Lefrançois, O., Matton, N., Causse, M., Gourinat, Y., 2016. The role of pilots' monitoring strategies in flight performance. In: Droog, A., Schwarz, M., Schmidt, R. (Eds.), Proceedings of the 32nd Conference of the European Association for Aviation Psychology, Groningen, NL.
Lim, B., 2014. Emerging threats from cyber security in aviation: challenges and mitigations. J. Aviat. Manag. Singapore 81–91.
Mancuso, V.F., Christensen, J.C., Cowley, J., Finomore, V., Gonzalez, C., Knott, B., 2014. Human factors in cyber warfare II. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 58 (1), 415–418. https://doi.org/10.1177/1541931214581085.
Manzey, D., Gerard, N., Wiczorek, R., 2014. Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. Ergonomics 57 (12), 1833–1855. https://doi.org/10.1080/00140139.2014.957732.
Moray, N., Inagaki, T., Itoh, M., 2000. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. J. Exp. Psychol. Appl. 6 (1), 44–58. https://doi.org/10.1037/1076-898X.6.1.44.
Mosier, K.L., Fischer, U.M., 2010. Judgment and decision making by individuals and teams: issues, models, and applications. Rev. Hum. Factors Ergon. 6 (1), 198–256. https://doi.org/10.1518/155723410X12849346788822.
Mosier, K., Fischer, U., Hoffman, R., Klein, G., 2018. Expert professional judgments and "naturalistic decision making". In: Ericsson, A. (Ed.), Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, Cambridge, MA (in press).
Muir, B.M., 1987. Trust between humans and machines, and the design of decision aids. Int. J. Man Mach. Stud. 27 (5–6), 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5.
Mumaw, R.J., 2017. Analysis of alerting system failures in commercial aviation accidents. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 61 (1), 110–114. https://doi.org/10.1177/1541931213601493.
Nguyen, D., Shelton, J.W., Mitchell, T.M., 2017. U.S. Patent No. US 9,836,990 B2. U.S. Patent and Trademark Office, Washington, DC.
Oakley, B., Mouloua, M., Hancock, P., 2003. Effects of automation reliability on human monitoring performance. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 47 (1), 188–190. https://doi.org/10.1177/154193120304700139.
Orasanu, J., Dismukes, R.K., Fischer, U., 1993. Decision errors in the cockpit. Proc. Hum. Factors Ergon. Soc. 37 (4), 363–367.
Orasanu, J., Fischer, U., 2014. Finding decisions in natural environments: the view from the cockpit. In: Zsambok, C.E., Klein, G. (Eds.), Expertise. Naturalistic Decision Making. Taylor and Francis, Hoboken.
Owens, W.A., Dam, K.W., Lin, H., 2009. Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities. National Academies Press, Washington, D.C.
Peryer, G., Noyes, J., Pleydell-Pearce, K., Lieven, N., 2005. Auditory alert characteristics: a survey of pilot views. Int. J. Aviat. Psychol. 15 (3), 233–250. https://doi.org/10.1207/s15327108ijap1503_2.
Piètre-Cambacédès, L., Chaudet, C., 2010. The SEMA referential framework: avoiding ambiguities in the terms "security" and "safety". Int. J. Crit. Infrastruct. Protect. 3 (2), 55–66. https://doi.org/10.1016/j.ijcip.2010.06.003.
Proctor, R.W., Chen, J., 2015. The role of human factors/ergonomics in the science of security: decision making and action selection in cyberspace. Hum. Factors 57 (5), 721–727. https://doi.org/10.1177/0018720815585906.
Rasmussen, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans. Systems Man Cybernetics, SMC 13 (3), 257–266. https://doi.org/10.1109/TSMC.1983.6313160.
Régis, N., Dehais, F., Rachelson, E., Thooris, C., Pizziol, S., Causse, M., Tessier, C., 2014. Formal detection of attentional tunneling in human operator–automation interactions. IEEE Trans. Hum. Machine Syst. 44 (3), 326–336. https://doi.org/10.1109/THMS.2014.2307258.
Rempel, J.K., Holmes, J.G., Zanna, M.P., 1985. Trust in close relationships. J. Pers. Soc. Psychol. 49 (1), 95–112. https://doi.org/10.1037/0022-3514.49.1.95.
Reynal, M., Colineaux, Y., Vernay, A., Dehais, F., 2016. Pilot flying vs. pilot monitoring during the approach phase. In: Boy, G.A. (Ed.), Proceedings of the HCI Aéro, pp. 1–7. https://doi.org/10.1145/2950112.2964583.
Rosekind, M.R., Gander, P.H., Miller, D.L., Gregory, K.B., Smith, R.M., Weldon, K.J., Lebacqz, J.V., 1994. Fatigue in operational settings: examples from the aviation environment. Hum. Factors J. Hum. Factors Ergon. Soc. 36 (2), 327–338. https://doi.org/10.1177/001872089403600212.
Sampigethaya, K., Poovendran, R., 2013. Aviation cyber–physical systems: foundations for future aircraft and air transport. Proc. IEEE 101 (8), 1834–1855. https://doi.org/10.1109/JPROC.2012.2235131.
Sarter, N.B., Mumaw, R.J., Wickens, C.D., 2007. Pilots' monitoring strategies and performance on automated flight decks: an empirical study combining behavioral and eye-tracking data. Hum. Factors 49 (3), 347–357. https://doi.org/10.1518/001872007X196685.
Schmider, E., Ziegler, M., Danay, E., Beyer, L., Bühner, M., 2010. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. Methodology 6 (4), 147–151. https://doi.org/10.1027/1614-2241/a000016.
Sorkin, R.D., 1988. Why are people turning off our alarms? J. Acoust. Soc. Am. 84 (3), 1107–1108. https://doi.org/10.1121/1.397232.
Spain, R.D., Bustamante, E.A., Bliss, J.P., 2008. Towards an empirically developed scale for system trust: take two. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 52 (19), 1335–1339. https://doi.org/10.1177/154193120805201907.
Starcke, K., Brand, M., 2016. Effects of stress on decisions under uncertainty: a meta-analysis. Psychol. Bull. 142 (9), 909–933. https://doi.org/10.1037/bul0000060.
Tessier, C., Dehais, F., 2012. Authority management and conflict solving in human-machine systems. Journal Aerospace Lab 4, 1–10.
Vereinigung Cockpit, 2017. SafeSKY 2017. Frankfurt.
Vries, P. de, Midden, C., Bouwhuis, D., 2003. The effects of errors on system trust, self-confidence, and the allocation of control in route planning. Int. J. Hum. Comput. Stud. 58 (6), 719–735. https://doi.org/10.1016/S1071-5819(03)00039-9.
Wickens, C.D., Rice, S., Keller, D., Hutchins, S., Hughes, J., Clayton, K., 2009. False alerts in air traffic control conflict alerting system: is there a "cry wolf" effect? Hum. Factors J. Hum. Factors Ergon. Soc. 51 (4), 446–462. https://doi.org/10.1177/0018720809344720.
Wilshusen, G.C., 2013. Cybersecurity: a Better Defined and Implemented National Strategy Is Needed to Address Persistent Challenges. Washington, D.C.
Yeh, M., Wickens, C.D., 2001. Display signaling in augmented reality: effects of cue reliability and image realism on attention allocation and trust calibration. Hum. Factors J. Hum. Factors Ergon. Soc. 43 (3), 355–365. https://doi.org/10.1518/001872001775898269.
Youden, W.J., 1950. Index for rating diagnostic tests. Cancer 3 (1), 32–35.
Zan, T., de, d'Amore, F., Di Camillo, F., 2016. The Defence of Civilian Air Traffic Systems from Cyber Threats. Rome.