Issue 25(2), 2025, pp. 97-121
ISSN: 1567-7141
TU Delft OPEN Publishing

EJTIR

Research article
https://ejtir.tudelft.nl
DOI: 10.59490/ejtir.2025.25.2.7370

Gamification in Automated Air Traffic Control: Increasing Vigilance Using Fictional Aircraft

Călin Andrei Badea[1], Gijs de Rooij[2], Clark Borst[3], Max Mulder[4]

[1] Control and Simulation, Faculty of Aerospace Engineering, TU Delft, The Netherlands; 0000-0002-5462-3559
[2] Control and Simulation, Faculty of Aerospace Engineering, TU Delft, The Netherlands; 0000-0003-4646-476X
[3] Control and Simulation, Faculty of Aerospace Engineering, TU Delft, The Netherlands; 0000-0001-9393-5304
[4] Control and Simulation, Faculty of Aerospace Engineering, TU Delft, The Netherlands; 0000-0002-0932-3979

Keywords
Air Traffic Control
Automation
Gamification
Vigilance

Publishing history
Submitted: 23 January 2024
Revised date(s): 12 February 2024, 2 January 2025
Accepted: 6 March 2025
Published: 18 June 2025

Cite as
Badea, C., de Rooij., G., Borst, C., & Mulder, M., (2025). Gamification in Automated Air Traffic Control: Increasing Vigilance Using Fictional Aircraft. *European Journal of Transport and Infrastructure Research, 25*(2), 97-121.

Abstract

The introduction of more advanced automation in air traffic control seems inevitable. Air traffic controllers will then take the role of automation supervisors, a role which is generally unsuitable for humans. Gamification, the use of game elements in non-gaming contexts, shows promising results in mitigating the effects of boredom in highly automated domains requiring human supervision. An example is luggage screening, where dangerous items are rarely found, through projecting fictional threats on top of x-ray scans. This paper presents and experimentally tests a proposed implementation of gamification within highly automated en-route air traffic control. Fictional flights were superimposed among automatically controlled real traffic, thus creating fictional conflicts that needed resolving. System supervisors were tasked to supervise the behaviour of a fully automated conflict detection and resolution system, while manually routing fictional flights safely and efficiently through the sector, avoiding conflicts with both real and fictional flights. Automation anomalies were simulated, as well as an automation failure event, after which the system supervisor needed to assume manual control over all traffic. The presence of fictional flights increased self-reported concentration levels and reduced boredom. However, some participants reported that fictional flights were distracting. Thus, while the use of fictional flights increases engagement, it might negatively affect other cognitive functions, and with that, compromise safety. Thus, while the implementation of such a tool might provide benefits in terms of skill retention and engagement, further research is recommended involving professional air traffic controllers, improved measurement tools and a longitudinal study that better excites boredom, complacency, and skill erosion in order to understand and mitigate its negative effects.

1 Introduction

The aviation industry has continually striven to maximise the efficiency and safety of commercial operations, driving the introduction of advanced technologies both in the cockpit and on the ground. The air traffic control (ATC) domain is predicted to undergo fundamental modernisation in the next 20 years, as the push for increased automation is gaining traction. While controller support tools, such as trajectory prediction and short-term conflict alerting (Noskievic and Kraus, 2017), are already widely used, the decision-making process still rests upon the air traffic controller (ATCO).

The SESAR Air Traffic Management Master Plan of 2020 (SESAR Joint Undertaking, 2019) aims for a high level of automation in ATC by the year 2040, facilitated by the development and widespread use of ADS-B technology. High levels of automation are expected to be first introduced at en-route air traffic service providers, whose ATCOs operate at a strategic timescale rather than at shorter timescale operations in approach or tower positions. An example of such a provider is Eurocontrol's Maastricht Upper Area Control Centre (MUAC), where fully automating (part of the) flights is currently being researched (Hendrickx and Tisza, 2019). The introduction of automatic conflict detection and resolution (CD&R) systems implies that ATCOs will undertake the role of supervisors, intervening only at the occurrence of exceptional events, such as automation failures. This will have major implications for the work environment and fundamentally changes the nature of the ATCO's task.

The Ironies of Automation (Bainbridge, 1983) describe potential issues that might arise when automating tasks previously performed by humans, while not excluding the human completely. Although much research has been conducted in the field of automation since, these issues have not yet been resolved (Strauch, 2018). Several negative effects of the human practice of highly monotonous supervisory tasks are described by Parasuraman et al. (2000). The required mental workload decreases when automation is introduced within a system, which leads to attention maintenance difficulties. Situation awareness is also affected, as humans are eliminated from the decision-making process. Several solutions to the cognitive issues posed by automation have been proposed and researched. Mercer et al. (Mercer et al., 2016) showed that simply involving controllers in the conflict detection task improves situation awareness, even if conflict resolution is fully automated. Borst et al. (Borst, Flach, and Ellerbroek, 2015) propose the use of ecological interfaces to help system supervisors detect faults in automation by increasing transparency and information flow within the human-machine system. Pop et al. (2012) s showed that producing and maintaining engagement improves failure detection in ATC supervisors, and is a key element in improving cognitive abilities.

However, one issue remains under-addressed: in a highly reliable and highly automated system, automation failures will be rare, and maintaining engagement in such an environment is more challenging when the intervention of operators is rarely required. Munyung et al. (2022) showed that keeping supervisors engaged in a highly automated ATC environment by asking task-related questions at regular time intervals can lower workload peaks experienced in case of automation failure, but does not improve subsequent manual control performance. One potential way to keep controllers both cognitively engaged and skilled is gamification, where game elements are used in non-gaming contexts (Zichermann and Cunningham, 2011). Threat image projection (TIP), a form of gamification, is a technique used in the highly monotonous task of airport luggage screening, and implies the superposition of fictional threats (such as firearms) onto luggage X-ray scans. Thus, the rate at which operators are exposed to potential threats is increased, thereby mitigating the effects of boredom (Cutler and Paddock, 2009).

This paper discusses an investigation of applying TIP within a highly automated en-route ATC environment through the use of fictional flights superimposed on real traffic, the latter being controlled by a fully automatic CD&R system. Fictional flights would require manual control (mitigating skill erosion) and the development and maintenance of a mental model of the flights in the sector (maintain situation awareness), thus inciting higher engagement while mitigating the effect of boredom. The increase in engagement is expected to have a positive effect on supervisory control performance, and seek to smooth the transition from supervisory to manual control in case of automation failure by maintaining the operator in the control loop. This type of tool could also be beneficial in low-workload situations (such as nighttime), helping maintain vigilance and focus within the work environment.

2 Background

2.1 The future of automation in ATC

As technological developments advance rapidly, the push for the introduction of automation in ATC is gaining traction. According to Nieto and Javier (2016), high levels of automation in ATC will result in increased safety and airspace capacity. However, this entails a progressive shift in the decision-making process from human operators to computers. Operators will assume the role of system supervisors (Ohneiser et al., 2018), a task that Foroughi et al. (2019) have experimentally determined to be unsuitable for humans.

The main source of human incompatibility with supervisory tasks stems from decreased mental workload, which leads to boredom. This deficiency is especially apparent if automation is highly reliable, as operator complacency and high trust can worsen supervisory control performance (Parasuraman, Molloy, and Singh, 1993). Situation awareness and vigilance are therefore negatively impacted, which leads to two types of human errors, as documented by Berberian et al. (2017) within the MINIMA project, which sought to create a measurement framework for vigilance:

- **Failure to detect:** leads to traffic supervisors missing automation failures or conflicts that might arise as a result of a malfunction;
- **Failure to understand:** leads to a lack of knowledge about the traffic situation and thus decreased performance whenever manual intervention may be required.

Berberian et al. (2017) mention the current proposed solutions for maintaining ATCOs in the loop in a highly automated ATC environment: human operator adaptation (training to prevent out-of-loop problems), system adaptations (dividing control tasks such that humans and machines perform suitable tasks), and adaptive automation (cooperation between operators and machines). However, the fourth solution that has the potential to facilitate the mitigation of out-of-the-loop issues is the concept of gamification.

De Rooij et al. (2024) propose automating only a part of air traffic, with part of them still requiring conventional manual control. This might have the same effect as fictional aircraft, but without introducing new elements to the work environment. However, this might result in mismatches between the actions performed by automation and the ATCO strategy, leading to unsafe situations or frustration.

One of the greatest barriers to increasing automation in the ATM domain is the low acceptance and scepticism among ATCOs for such tools or implementations. Multiple studies and surveys conducted with ATCOs (Bekier, Molesworth, and Williamson, 2012; Svensson et al., 2021; Langford et al., 2022) have consistently revealed a prevalent negative attitude towards full automation. In this context, full automation refers to a scenario in which the majority of decision-making responsibilities are delegated to automated systems, with ATCOs primarily serving in a supervisory capacity. This resistance stems from various factors, including concerns about job security, reduced situation awareness, and potential safety risks.

However, ATCOs demonstrate a more positive attitude towards assistive technologies designed to enhance situation awareness. These tools, which augment rather than replace human decision-making, tend to gain higher acceptance rates. Moreover, research indicates that, as ATCOs gain more exposure to such assistive technologies, their acceptance of these tools generally increases (Bekier and Molesworth, 2017). This suggests that a gradual introduction of automation, coupled with proper training and familiarisation, could be a more effective strategy for implementation.

An important factor in increasing trust and acceptance of automated systems among ATCOs is automation transparency (Hurter et al., 2022; Loft et al., 2023; Sargent and Wickens, 2023). Systems that perform decisions and actions with a high degree of transparency help ATCOs understand and predict system behaviour. This reduces the need for constant verification, potentially decreasing cognitive workload and increasing overall efficiency.

Another significant concern is the potential for skill degradation among ATCOs (Volz and Dorneich, 2020; Rogosic et al., 2021). As automated systems take over more tasks, there is a risk that controllers may lose proficiency in skills that are essential during system failures or unexpected situations. Strategies exist to mitigate the effects of skill degradation, including implementing more rigorous and frequent training programs that focus on maintaining essential skills, even as automation increases. Additionally, fostering effective teamwork and communication among ATCOs can help maintain situation awareness and decision-making abilities (Smith and Baumann, 2020).

2.2 *Concept of gamification*

Gamification is a technique that uses game elements in non-gaming contexts with the purpose of obtaining associated cognitive benefits. Elements such as leader-boards, scores and achievements can increase productivity and motivation by creating a sense of progression and reward (Zichermann and Cunningham, 2011). Gamification is currently used in a wide range of domains, such as management techniques (PACAS project (Paja et al., 2018), which proposes a change in the management style of ATM) or research (the use of mobile games for the simulation of HIV behaviours (Khatib et al., 2011)). The latter makes use of a gamification technique through which fictional goals, which are more appealing and motivating to the public (completing puzzles), are indirectly used to fulfil real goals (HIV simulation).

One successful implementation of gamification within the aviation industry is found in airport security luggage screening. Operators must supervise the flow of luggage through an X-ray scanning machine and signal the presence of prohibited items. These exceptional events rarely occur. Thus, the activity is considered to be highly monotonous, which can lead to a degradation in cognitive factors such as vigilance and situation awareness. Meuter and Lacherez (2016) proved that long shifts lead to an increase in threat detection errors. To mitigate this, some airports have implemented threat image projection (TIP), described by Schwaninger (2006) as the projection of fictional images of threats on X-ray machine screens, thus achieving an increase in the threat rate. The projected images cover a wide range of prohibited items, such as firearms or other weapons, and are superimposed over regular luggage scans. This superposition is also varied in its complexity and level of obstruction, achieving a wide range of viewing difficulty and complexity. According to Hofer and Schwaninger (2005), the increase in threat rate in combination with such a diverse fictional object database leads to better detection performance.

Compared to other types of gamification, TIP is implemented in a safety-critical environment, facilitating the porting of the concept to ATC. The gamification elements and concepts used within TIP have the potential to increase supervisors' engagement within the automated work environment. It can be used to provide operators with a secondary task that requires them to interact with elements on the screen, which could lead to a deeper understanding of the situation, as well as better automation fault detection rates.

2.3 *Ethical considerations*

The concept of gamification has attracted criticism from an ethical standpoint due to its nature and proposed implementations. By considering gamification in an ATC context, two ethical considerations must be considered when designing an implementation. Firstly, according to Kim and Werbach (2016), gamification can create a manipulative environment in which participants are forced to partake in a game without active consent. This is apparent when considering TIP, as screening operators do not have a choice when it comes to participation, and are not aware in advance of which threats are fabricated.

Secondly, such implementations of gamification can also produce unnecessary stress to participants, as the stakes are artificially raised (Kim and Werbach, 2016). In TIP, the operators do not have prior knowledge about the nature of a threat, and are only told whether it is fictional or not after they have identified it. This can have two negative effects: on the one hand it increases the stress level of operators, on the other hand, repeated fictional threats can lead to a desensitisation towards real threats.

These issues were considered for the design of the tool proposed in this work by maximising its transparency and ensuring that users are aware of the fictional nature of the threats. Furthermore, although not implemented within the experiment, a potential future implementation could allow users to toggle this tool to their liking.

3 Design and Implementation

The main feature of TIP is that fictional threats are introduced to the work space of a system supervisor. Within the ATC domain, the most prominent threat to the overall safety level of operations is aircraft conflicts. The proposed implementation superimposes fictional flights onto real traffic to create virtual conflicts that require operator intervention, evoking a higher level of engagement when compared to a purely supervisory task. Unlike real flights, which would all be controlled by automation, fictional flights must be manually commanded by system supervisors.

3.1 Control task analysis

The intended effect of gamification is to maintain the system supervisor within the control task loop. Figure 1 is a representation of the control task based on the decision ladder diagram used by Borst et al. (2019) to define the steps an operator takes from realising the need of intervention to the execution of a command. In the event of automation failure, if the system supervisor experiences low vigilance and situation awareness, the decision ladder is traversed fully, starting from entry point **A**. Observation and identification are required to understand the current state of traffic and formulate a command. Fictional flights, however, can increase situation awareness before the failure event, enabling shortcuts to points **B** or even **C** when responding to the failure. This could lead to a better performance when transitioning from supervisory to manual control.

With the introduction of fictional aircraft, the control task of operators can be divided into two types: supervisory control and manual control. While performing supervisory control, the operator supervises the fully automated CD&R system with the help of ATC tools (e.g., short term collision avoidance alerts) while also performing the task of manually controlling fictional aircraft. This secondary task requires the operator to exercise manual control skills and to build a mental model encompassing both real and fictional aircraft, as it potentially introduces `mixed conflicts' between fictional and real aircraft. The second type of control task that operators must perform in the case of automation failure is the manual control of real aircraft, which does not differ from the current non-automated ATC task. Fictional aircraft should thus not be a part of this control task in order to avoid increasing the workload over a manageable threshold for controllers.

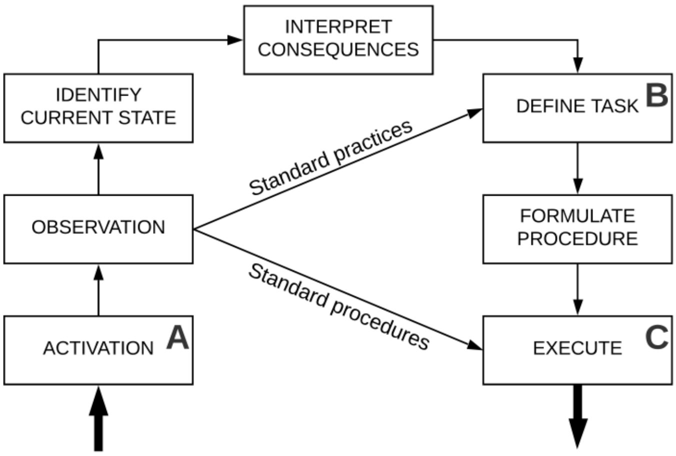


Figure 1. *Simplified decision ladder diagram (Borst, et al., 2019).* ***A*** *is the standard entry point when transitioning from manual to supervisory control. If the human operator is kept within the control loop, shortcuts after activation to* ***B*** *or* ***C*** *are more accessible.*

3.2 *Work space integration*

Fictional flights were implemented such that operators would be aware of their nature and that they were distinguishable from real flights. This increases transparency compared to TIP, where operators are not initially aware of the fictional nature of threats. It also facilitates the differential prioritisation of tasks. However, it might lower vigilance, as controllers will assign lower priority to fictional flights.

The design presented in Figure 2 was created by considering the display design principles of discriminability and redundancy described by Wickens et al. (2004),. Fictional flights differ from real flights in both colour and shape. The colour blue was chosen due to its high contrast with the dark backgrounds often used in ATC displays. In addition, blue is currently used at MUAC to indicate flights that are within an ATCO's airspace, but controlled by an ATCO of an adjacent sector.

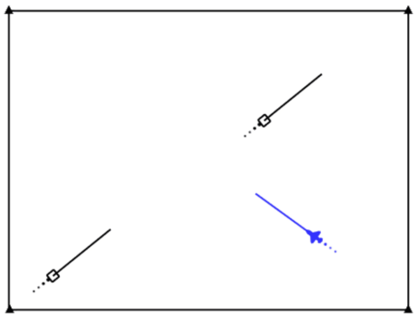


Figure 2. *Visualisation of a fictional flight introduced on the ATC radar screen.*

By analysing the design considerations of TIP (Schwaninger, 2006), two main factors that affect the influence of fictional flights on operators can be determined: the complexity of the traffic situation and the number of threats present on the screen. The first can be correlated with the nature of the conflicts induced by fictional flights (e.g., head-on, catching up) which can be influenced by their traffic pattern. If an airspace sector is assumed to contain several main routes along which most flights will travel, there are three types of traffic patterns relative to these routes: (1) fictional flights are introduced such that they cross the main traffic flow, (2) fictional flights are introduced among the flights of the main flow, or (3) as a combination of among and crossing the main flow. The second influencing factor is correlated with the number of fictional flights present in the sector at one time, which can influence the workload and concentration that an operator must allocate towards them. A high number of fictional flights demands a higher workload, thus making the supervisory task more difficult to perform. This is undesirable, as supervising air traffic is the highest priority.

It is expected that the presence of fictional flights will always lead to higher workload when compared to an unaided supervisory control situation. However, if all fictional flights are found within the main traffic flow, the workload required is lower than in the other situations, as they will create little conflict with the flights around them. From a situation awareness perspective, a combined fictional flight traffic pattern produces more evenly distributed traffic in the sector, thus aiding the operator in maintaining an overview of all parts of the sector. Therefore, introducing a low number of fictional flights, both within and across the traffic flows is hypothesised to yield the best results.

4 Method and Experimental Setup

As gamification has not been previously used in the context of ATC in the form presented in this paper, an exploratory experiment has been conducted to obtain more insight into the effects of fictional flights on the cognitive performance of controllers. The goal of the experiment was not only to determine whether the performance of participants was affected, but also to gather subjective feedback from peers with various backgrounds and previous ATC experience.

The experiment was designed for a controlled and simplified ATC environment. This was done due to the participant pool being selected from among faculty members who had previous experience with ATC experiments but mostly did not undergo professional ATC training. The participants assumed the role of a system supervisor and were given the task of reporting anomalous events that might occur when automation is in charge of controlling flights. Participants would also have to intervene when automation would experience a failure, thus transitioning from supervisory to manual control of flights. Thus, the experiment scenario run was divided into two phases: a supervisory control phase and, after the failure event, a manual control phase.

4.1 Apparatus and software

The experiment was conducted using SectorX, a TU Delft in-house developed Java-based ATC simulator, which was modified to include fictional flights, a supervisory control mode and a manual control mode. The hardware setup is presented in Figure 3. The screen was oriented such that no distracting elements were in view, and SectorX was run in full screen mode to hide the task bar and clock of the operating system. Separation circles with a radius of 2.5NM, history dots, and one minute look-ahead velocity vectors were added to the aircraft blips to facilitate the supervision task.

Figure 3. The hardware setup used for the experiment.

During the supervisory phase of a scenario, automation was enabled and handled all real flights, while only fictional flights, if present, could be manually controlled. The manual control phase began when the scenario automation failure time was reached. Fictional flights disappeared from the screen, and the manual control of the real flights was enabled.

The simulator was augmented to accommodate the anomaly reporting during the supervisory control phase. While automation was active, unusual behaviour of real flights could be reported by clicking on the offending flight and typing a report in a pop-up window, as shown in Figure 4. When automation failed, a notification appeared on screen, (Figure 5) which needed to be acknowledged and dismissed by clicking on it, after which the ability to manually cotntrol flights was enabled.

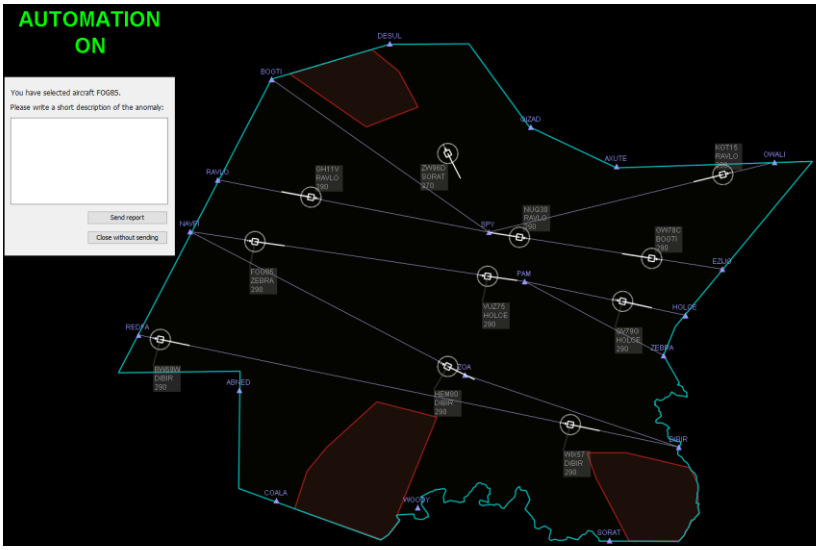


Figure 4. Simulator screen during the supervisory control phase, with anomaly reporting box present.

Figure 5. Simulator screen in case of automation failure, with alert notification present.

4.2 Participants

Sixteen participants volunteered in the experiment. Most had previous ATC experience through the means of university courses as well as previous experience with the SectorX ATC simulator. A between-participants experiment design was used, with participants assigned to either a fictional flights group or a baseline group. As the volunteers originated from different academic levels within the faculty of Aerospace Engineering of TU Delft, the distribution presented in Table 1 was made to balance the groups as much as possible.

Table 1. Participant characteristics.

Group:	Fictional flights group	Control group
Number of participants	8	8
MSc students	5	4
PhD students	2	3
Academic staff	1	1

4.3 Participant tasks

Participants were instructed to perform the following tasks:

- **Primary supervisory control phase:** supervise traffic controlled by automation, and report anomalies when predicted to occur. Write a short description of the anomaly and send the report. Avoid false-positive reporting.
- **Secondary supervisory control phase (fictional flights group only):** route fictional flights safely and efficiently towards their exit waypoints. Avoid conflicts with real and other fictional flights. Automation does not account for the presence of fictional flights, thus compensate and command fictional flights in case a conflict arises as a result of an automatic aircraft manoeuvre.
- **Primary manual control phase:** after the automation failure, dismiss the notification as soon as possible and proceed with routing flights towards their exit waypoints safely and efficiently until the end of the traffic scenario. Automation will not re-enable.

4.4 Air traffic scenarios

A clipped version of the Delta sector of MUAC's en-route airspace was selected for the experiment. ADS-B data from the year 2018 were analysed and used to develop realistic traffic patterns, as shown in Figure 6.

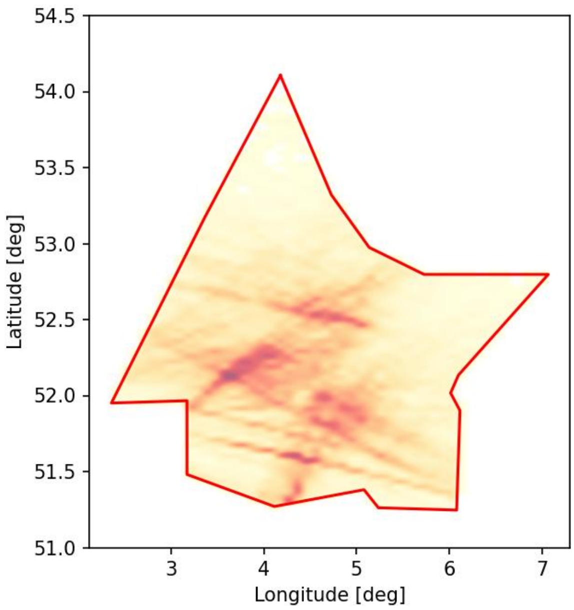


Figure 6. Heat map of flight routes extracted from ADS-B data used to generate traffic patterns.

Three restricted airspace areas were placed at the edges of the sector to test the situation awareness of participants with events occurring away from the centre of the screen. Entry and exit waypoints were distributed along the sector boundary, as well as three inner waypoints coinciding with the high traffic areas. The final sector configuration is presented in Figure 7.

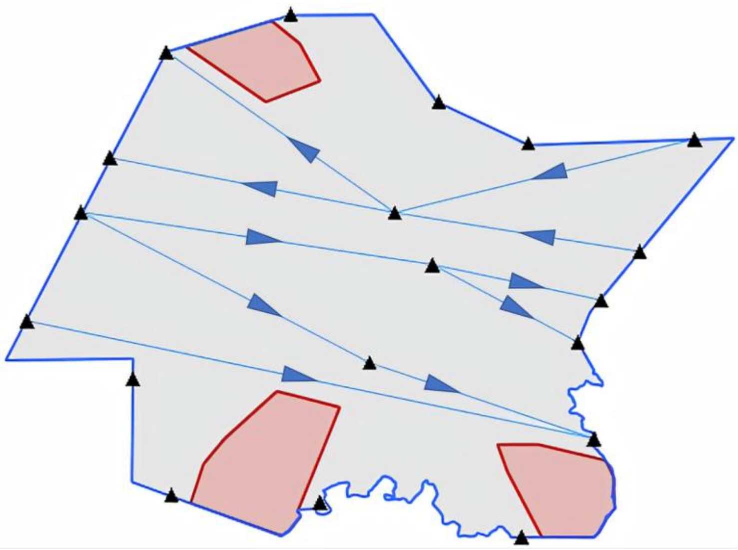


Figure 7. Modified Delta sector used for the experiment traffic scenario.

The baseline scenario had a traffic density of 45 flights per hour, extracted from the aforementioned ADS-B data, with approximately eleven aircraft within the sector at all times (see Figure 3). The length of the scenario was based on the attention decrement phenomenon described by Mackworth (1948), who experimentally showed that the greatest decrement in attention while performing a monotonous supervisory task occurs in the first hour. Hancock (2012) expands on the concept and links the attention decrement to the nature of the task. Based on these considerations, the length of the supervisory control part of the scenario was set at 4,000 seconds (ca. 67 minutes) and the length of the manual control part at 700 seconds (ca. 12 minutes).

The fictional flights scenario was built by adding fictional flights to the baseline scenario both within and crossing the main traffic flows. The real flights were identical between the two groups. Through several iterations and preliminary test participants, it was decided that fictional flights would represent 20% of the real flights on screen, therefore approximately two fictional flights would be present in the sector at all times.

4.5 Automation, anomalies and failure

Traffic scenarios included scripted commands such that real flights would maintain a separation of approximately 7NM from each other, follow the routes presented in Figure 7, and manoeuvre at the three interior waypoints when possible to increase automation transparency and predictability. As the commands were scripted, the presence of fictional flights had no impact on them, thus conflicts could arise between fictional and real flights that needed human intervention. While a command was executed, the labels of the manoeuvring aircraft would change in colour. No other automated aiding tools (such as conflict detection) were available during the experiment.

Three types of anomalous events were triggered throughout the supervisory phase of the scenario, selected for the objectivity with which they could be spotted:

- **Loss of separation:** two flights breach the minimum lateral separation requirement of 5 NM;

- **Restricted airspace separation violation:** flights get close (less than 2.5 NM) or breach restricted airspace areas;
- **Incorrect exit waypoint:** flights exit the sector through a different waypoint than assigned.

In total, seven anomalous events occurred at the times listed in Table 2. After 3,015s, two anomalies occurred simultaneously in different regions of the screen to provide insight into the occurrence of attention tunnelling (i.e., attention is drawn towards one part of the screen).

Table 2. Anomalous events within the experiment scenarios.

Time [s]	Anomaly type
797	Restricted airspace violation
1,603	Incorrect exit waypoint
2,033	Loss of separation
2,520	Restricted airspace violation
3,015	Restricted airspace violation
3,015	Incorrect exit waypoint
3,870	Restricted airspace violation

Lastly, a total automation failure occurred after 4,000s, to test the performance of operators transitioning from supervisory control to full manual control. An alert message was shown on the screen, which the operator had to dismiss by clicking on it (Figure 5). From then on, the operator had to manually control all flights in the sector, with no elements of automation enabled (i.e., no conflict detection or other aiding tools). Any present fictional flights were removed from the screen, to let the participants dedicate all their resources towards the manual control of real flights, while also levelling the conditions in which manual control performance was measured. A visual summary of the scenarios is presented in Figure 8.

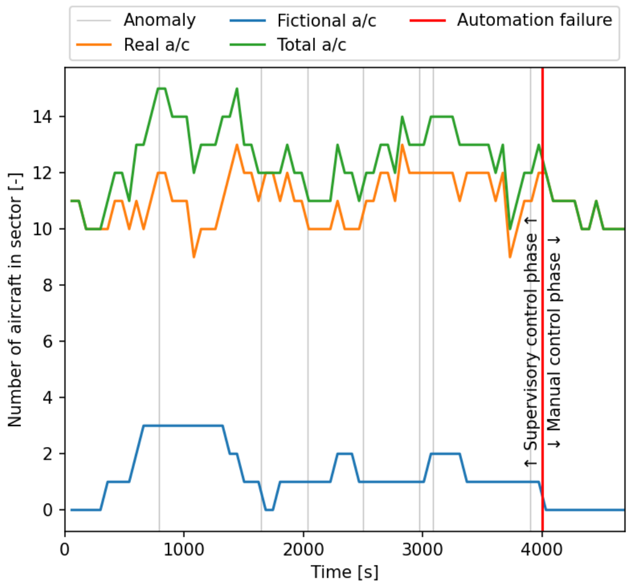


Figure 8. Overview of a typical traffic scenario. The number of aircraft over time will vary depending on the control performance of participants.

4.6 Experiment procedure

The procedure used for each participant is presented in Figure 9. After receiving a short briefing, participants underwent a series of training scenarios, both with and without fictional flights before being informed of their group assignment. Thus, all participants were introduced to, and experienced the presence of a fictional aircraft within the ATC environment.

Within the experiment itself, the supervisory control phase differed between the groups. The baseline group was tasked with supervising an automated traffic management system controlling real aircraft, with no fictional aircraft present. The fictional aircraft group had the same task, but also had to route fictional aircraft through the airspace. After an automation failure that marked the end of the supervisory phase, fictional flights (if present) would disappear, and both groups had to manually control the remaining real flights for another 12 minutes. The experiment was concluded with a survey.

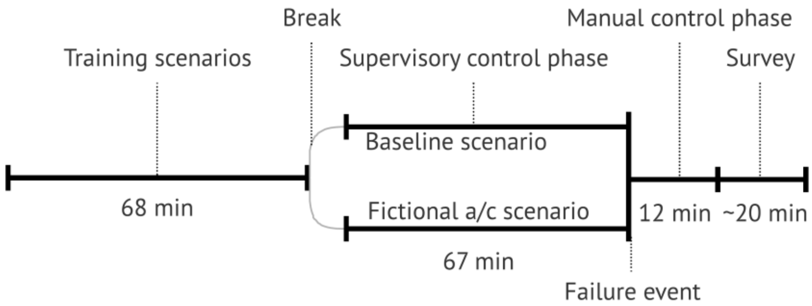


Figure 9. Experiment timeline.

4.7 Control variables

Due to the novelty of the concept, as well as the expected high variability in ATC characteristics and experience among participants, the following variables were controlled during the experiment:

- **Real traffic and scripted commands:** identical for all participants during both supervisory and manual control phases.
- **Degrees of freedom:** aircraft control was limited to heading commands. All aircraft flew at FL290 with constant indicated airspeeds between 250 and 310 kts.
- **Anomalous automation events:** identical across the fictional flights and baseline scenarios.
- **Radar update rate:** once every 5 seconds (0.2 Hz).

4.8 Dependent measures

- **Supervisory control performance**
  - o *Anomaly reports:* time and description.
  - o *Alert reaction time:* time elapsed between the moment the alert is shown on screen and the click that dismisses it.
- **Manual control performance**
  - o *Clearances:* heading commands issued by participants.
  - o *Average heading deviation:* per flight in degrees, which increases if flights are not following the ideal path towards their destination.
- **Engagement**
  - o *Mouse clicks:* including clicks on flights or their labels.
  - o *Mouse click rate ratio:* ratio between the average click rate before and after automation failure per Equation 1.

$$CRR = \frac{Number\ of\ clicks\ after\ failure\ /\ 700}{Number\ of\ clicks\ before\ failure\ /\ 4000} \tag{1}$$

- **Subjective questionnaire**
  - o *Situation awareness:* using SART questions (EUROCONTROL, 2012), a method to assess the ability of participants to sustain attention over a given time period.
  - o Several Likert-scale and open questions about control strategies, order of priorities when supervising automation, experience with and trust in automation and experience with fictional flights, if present.

4.9 Hypotheses

It was hypothesised that superimposing fictional flights on real flights in a highly automated ATC environment would improve the operator's:

HP-1 anomaly detection rates (i.e., minimise detection misses),
HP-2 vigilance levels,
HP-3 manual control performance (safety and efficiency indicators) immediately after automation failure, and
HP-4 situation awareness.

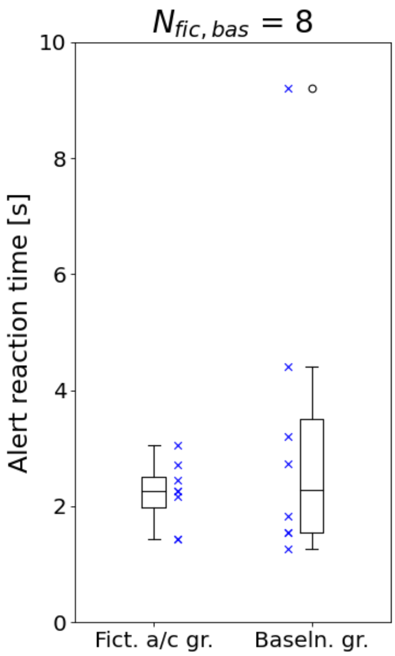
5 Results

This section presents the objective measures in terms of supervisory and manual control performance, engagement and subjective questionnaire results. Due to the small sample size (eight per group), the analysis was mainly focused on qualitative observations, with Mann-Whitney U statistical tests presented where feasible.

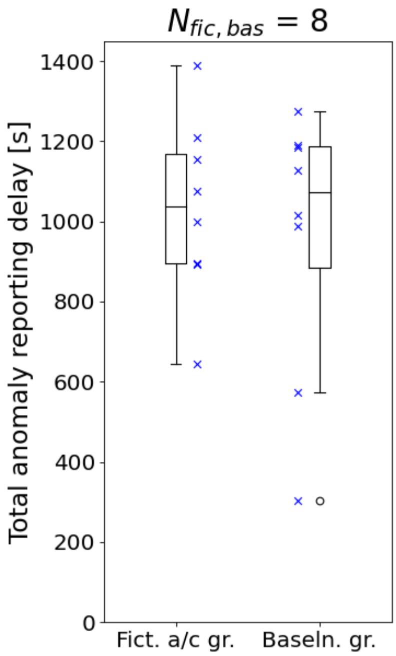
5.1 *Supervisory control performance*

Starting with supervisory performance, Figure 10a portrays the reaction times of participants to the automation failure notification. While the medians for the two groups are identical, the baseline group's reaction time shows a larger spread. The outlier within the baseline group is a participant who inadvertently attempted to switch to manual control without first dismissing the alert. No significant difference between the two groups is observed.

The same remark can be made when considering the total anomaly reporting delay per participant during the supervisory phase, presented in Figure 10b. The variance of the dataset is relatively large due to the diversity in participant strategy when reporting anomalies, which they were told to report only when being confident about them.



(a) Automation failure alert reaction time from appearance on screen until participant dismissal



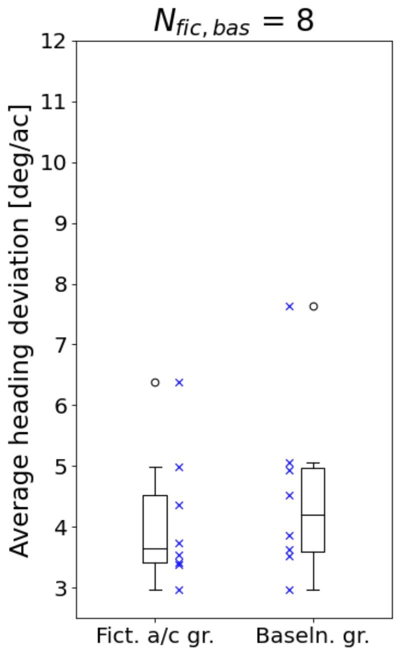
(b) Total reporting delay with respect to the fastest report time for each anomalous event

Figure 10. Supervisory performance per experiment group.

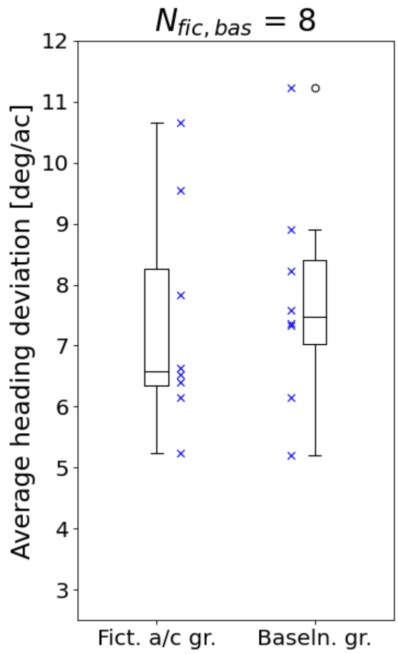
It should be mentioned that all participants detected all anomalies that occurred during the scenario. However, a notable result can be observed when considering false positive anomaly reports (mostly consisting in flights manoeuvring late): most of these were submitted by participants in the fictional flights group (5/8 participants), whereas only one baseline group participant submitted false positive reports.

5.2 *Manual control performance*

The average heading deviation does not reveal a significant difference either, although the median of the fictional flights group is higher both over the entire time interval after automation failure (Figure 11a) and within 2 minutes after failure (Figure 11b). A relatively high variability in the data set can also be observed, as the participants used different strategies when manually controlling flights. It should be noted that the heading coefficient does not capture all aspects of performance, and in essence represents the time efficiency with which participants solved the immediate conflicts after failure.



(a) Throughout manual control



(b) First 120 seconds after failure

Figure 11. Manual control performance (average heading deviation) per group.

A noteworthy trend can be observed in the number of commands issued to aircraft by each participant, presented in Figure 12. No significant difference can be observed when comparing the number of commands issued to real aircraft between the two groups. However, within the fictional aircraft group (left side), the number of commands issued to real aircraft and fictional aircraft is correlated for each participant. This indicates that both types of aircraft were treated similarly by the participants in this group.

5.3 *Engagement*

In terms of engagement, a significant difference between the experimental groups was observed in the ratio between the label click rate before and after automation failure (N=8, U=12, p=.038). From Figure 13a, it can be seen that the fictional flights group has a much lower variability in the label click rate ratio. While the absolute number of clicks is a matter of personal strategy, the {\em ratio} between the click rates is an indicator of the consistency with which participants interacted with the labels throughout the experiment. The mean ratio of around 1 for the fictional flights group indicates that this group was more consistent in interacting with labels, whereas the baseline group had a higher click rate after automation failure than before. A similar significant trend (N=8, U=10, p=.021) is seen when considering all mouse clicks (Figure 13b). The more consistent clicking strategy in the fictional flights group suggests a steadier transition from supervisory to manual control.

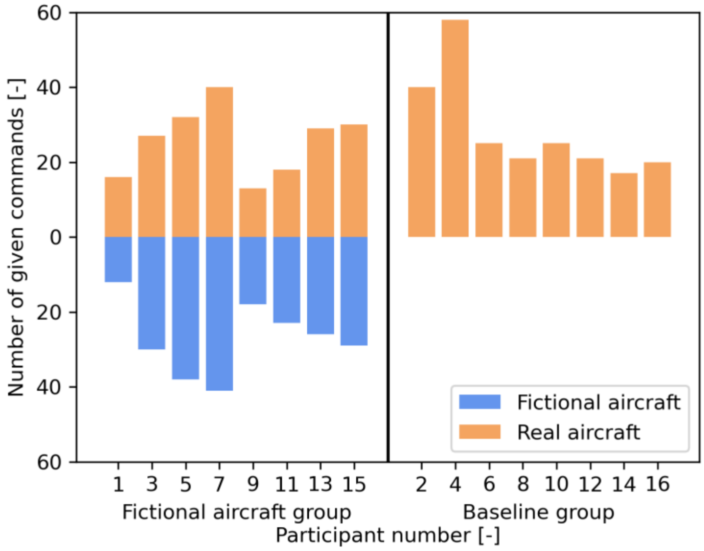
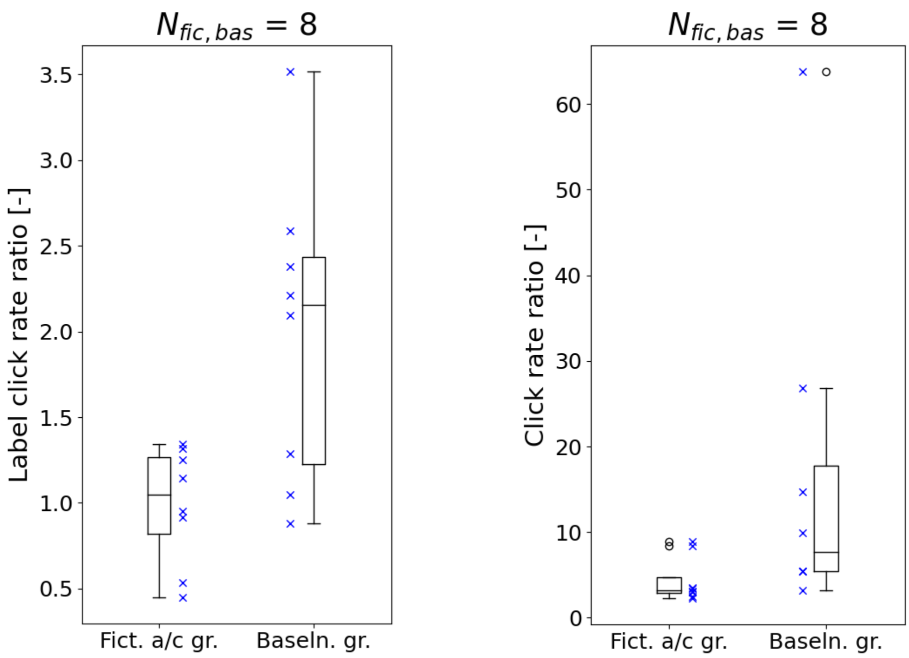


Figure 12. *Number of commands given to aircraft by each participant. The left half of the figure represents the fictional aircraft group, that issued commands to both real and fictional aircraft. As fictional aircraft were not included in the experiments of the baseline group, presented in the right half, the number of commands issued to fictional aircraft is null.*



(a) Clicks on flight labels (b) All clicks

Figure 13. *Mouse click rate ratios split per group.*

The difference in click rates between the two groups can also be seen when the cumulative number of mouse clicks is plotted over time (Figure 14). Due to technical difficulties, this timestamped click data was only recorded for five participants in each group. The activity of the participants is comparable during the first 1,000 seconds of the simulation, when at most one fictional flight is present. In line with the initial peak of fictional flights entering the sector between 1,000 and 2,000 seconds (Figure 8), the activity of the fictional flights group is higher during this period as interactions between fictional and real flights become more apparent. However, shortly after the failure event (between 4,000 and 4,200 seconds), the average click rate of the baseline group is higher, suggesting that the change in mouse activity when transitioning to manual control is more sudden among this group.

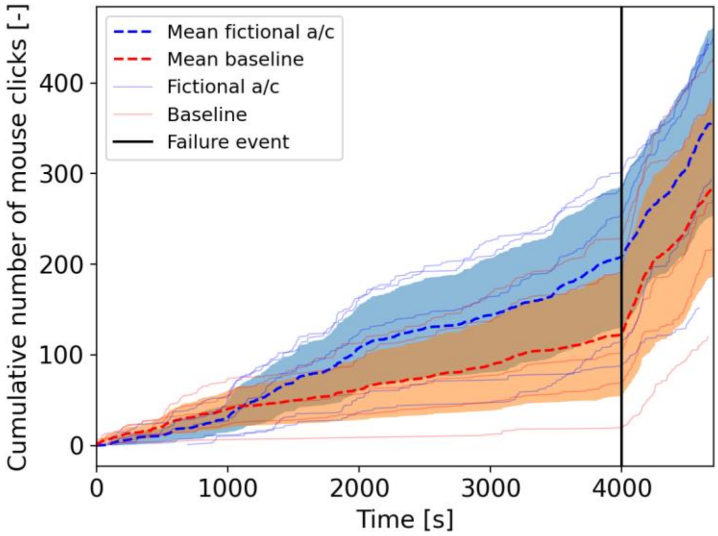


Figure 14. Cumulative mouse clicks over time for five participants per group. Also shown are the averages per group and their 95% confidence intervals.

This is also reflected in the number of commands issued to aircraft after the failure event, presented in Figure 15. Overall, participants in the baseline group gave more commands within the first 200 seconds after the failure event when compared to the fictional aircraft group.

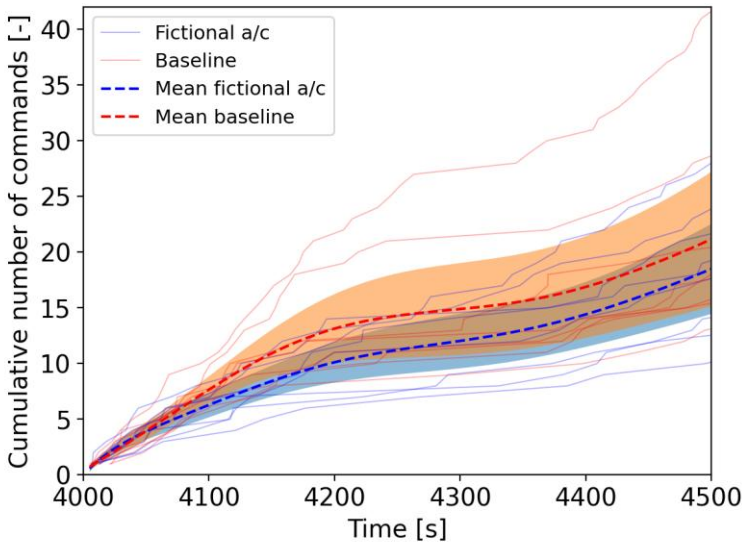


Figure 15. Cumulative number of commands per participant in the manual control phase, as well as means per group with a 95% confidence interval.

5.4 *Subjective questionnaire data*

This section presents the subjective data obtained from the survey at the end of the experiment, focusing on the questions that the participants had to answer using a 1-7 Likert scale. The SART index computed from the answers did not yield a significant difference between the fictional flights and the baseline groups, as presented in Figure 17. The results for this metric are also highly variable, indicating that the traffic scenarios of the experiment were perceived differently by the participants.

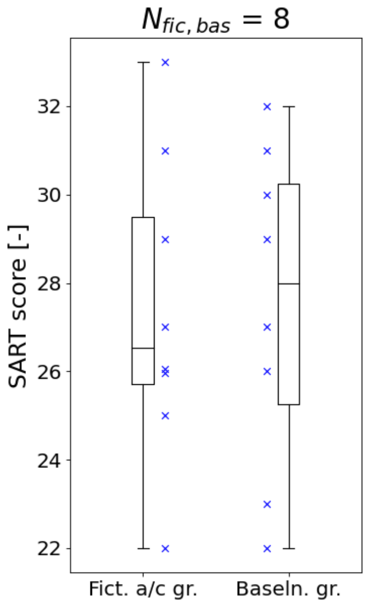


Figure 16. Results of the SART computation for both experimental groups.

However, notable differences can be observed among the answers to the individual SART questions, presented in Figure 17. The concentration level ratings (*How much did you have to concentrate on the situation? Were you experiencing full concentration (7) or little to no concentration at all (1)?*), show that participants in the fictional aircraft group had to concentrate more than the participants in the baseline group, which was one of the goals of the proposed tool (i.e., to increase required concentration and thus workload).

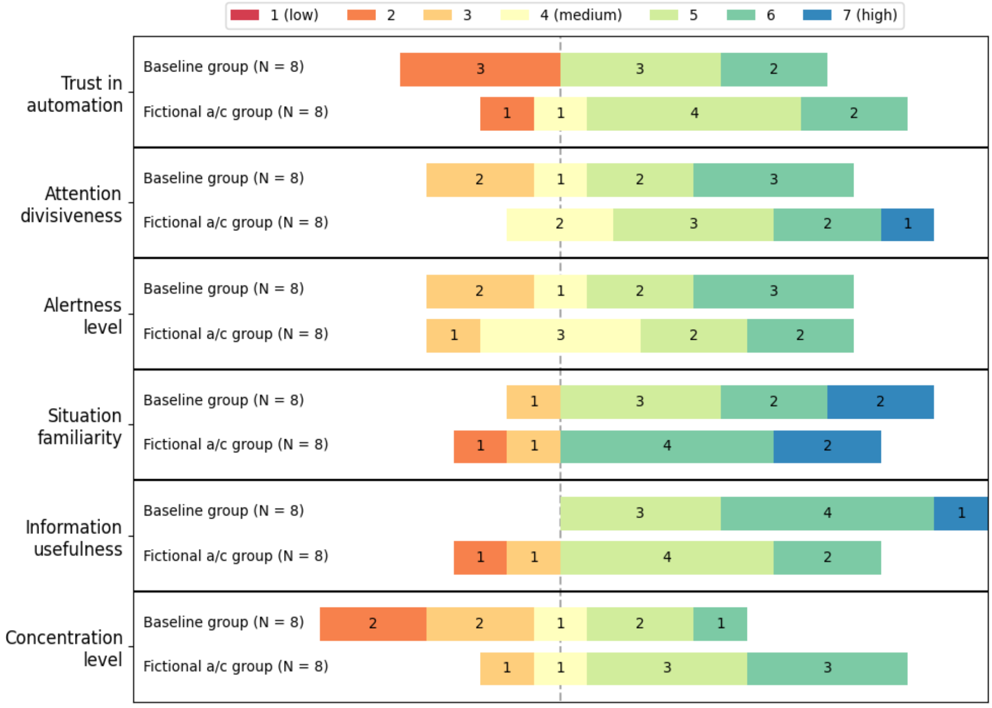


Figure 17. Answers to selected Likert-scale SART questions.

The answers to the information usefulness question (*How much did the received information help with understanding the situation? Have you understood a great deal of knowledge (7) or very little (1)?*) reveal another trend: participants in the baseline group indicated that they had a greater understanding of the knowledge received from the display, whereas the fictional flights group scored lower overall.

A difference between the two groups can also be observed in the answers to the attention division level question (*How much was your attention divided in the situation? Were you focusing on many aspects of the situation (7) or focused on only one (1)?*): participants in the fictional flights group reported higher levels, which is expected due to there being more elements on the screen that require their focus. This, together with the answers to the informational usefulness question, indicates that the presence of fictional aircraft acted as a distracting element for some participants.

Lastly, participants were asked whether they found or would have found fictional aircraft beneficial for maintaining vigilance within the experiment environment. For the fictional aircraft group, 5/8 participants responded affirmatively, and 6/8 for the baseline group. Overall, more than half of the participants in both groups considered that fictional flights were or would have been useful.

5.5 Survey open questions

During the survey conducted at the end of the experiment, several open questions were asked with the purpose of gaining more insight into the experience and strategy of each participant. An important factor for the concept of fictional aircraft is the way in which participants treated the fictional threats, and whether this hindered their supervisory control performance. From the open questions, the fictional aircraft group participants reported two strategies for treating fictional aircraft as: (1) real aircraft, and (2) as low priority aircraft.

The effect of the treatment of fictional aircraft by participants can be seen in the reaction time for Anomaly3 (after ca. 34minutes), due to it being the only loss of separation event occurring while automation was active. The anomaly can be predicted or noticed for a shorter period of time compared to the other anomalies, and is the anomaly with the least variability among participants. For this reason, when grouping fictional aircraft participants according to their treatment of fictional aircraft, a pattern can be noticed in the Anomaly 3 reporting times, as shown in Figure 18: lower priority given to fictional aircraft by participants resulted in faster report times for Anomaly 3. This means that participants who did not differentiate between fictional and real aircraft in terms of priority performed worse in terms of supervisory control performance. On the other hand, the group that treated fictional aircraft as real aircraft did perform similarly to the baseline group, therefore it cannot be concluded that fictional aircraft resulted in lower performance than the baseline case when treated as real aircraft.

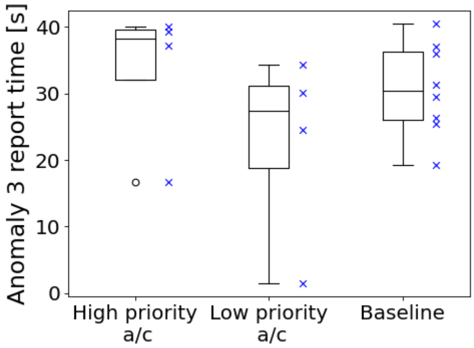


Figure 18. Anomaly 3 reporting time data with fictional aircraft participants divided in three groups depending on their treatment of fictional aircraft: as real, high priority aircraft (N = 4) and as low priority aircraft (N = 4). Baseline group added for comparison.

In the open question survey, participants in both groups reported the fact that anomalies helped with maintaining vigilance during the experiment. Thus, even though most participants indicated a high level of boredom, the overall decrease in cognitive performance was not significant enough to produce a visible effect in supervisory and manual control performance between the two groups.

6 Discussion

The work presented in this paper aimed to investigate the effects of using gamification within a highly automated ATC environment to enhance controller cognitive abilities when supervising automation. The implementation of gamification was made using fictional flights overlaid on the radar screen among automatically controlled real flights.

An analysis of the performance measures does not reveal significant differences between the fictional flights and the baseline groups due to the diversity in strategies and techniques between participants in both supervisory and manual control. Repeating this study with a sample of professional ATCOs is therefore recommended, as they are expected to have more consistent strategies for handling traffic.

The threshold required for a significant attention and performance decrement was not attained, as all participants reported all anomalies presented in Table 2 correctly. Contributing to this was the decision to make the types of anomalies known to participants beforehand. On the one hand, it provided more control over the experiment by reducing the ambiguity and confusion on what events can be considered anomalies. On the other hand, more extensive training could have achieved the same result, thus making the anomalies more difficult to spot while also reducing confusion. This is confirmed within the open questions answered by participants after the experiment: most participants from both groups reported that the occurrence of anomalies, as well as the prior knowledge of their existence, contributed positively towards maintaining vigilance, which enabled better anomaly detection performance. Thus, no conclusion can be drawn relating to hypotheses **HP-1** (fictional flights use improves anomaly detection rates) and **HP-2** (fictional flights use improves vigilance and reaction times), as no detection misses occurred. Moreover, the manual control performance measures (Figures 11a and 11b) show a high degree of variability with no significant difference between groups. Thus, conclusions cannot be drawn regarding hypothesis **HP-3** (presence of fictional flights improves manual control performance after an automation failure).

Experimental results show that the use of fictional flights can enhance some cognitive processes, and increase required concentration levels. Fictional flights also helped participants achieve more consistent engagement with the simulation: most participants in the baseline group were significantly more actively clicking after the failure event compared to before, while participants in the fictional flights group showed a more constant level of interactivity. This shows that, from the point of view of engagement, participants in the fictional flights group experienced a less sudden transition when changing from supervisory to manual control. However, this did not translate into a significant difference in manual control performance between the two groups in immediately after the failure event, which was mostly dependent on the personal control strategy of the participants.

The false-positive reports submitted by some participants provide some insight into the effects of fictional aircraft presence. While not being considered as anomalies, these events require the attention and supervision by the participants during the scenario. Thus, false-positive reports could be an effect of an overall increase in engagement and situation awareness achieved by the presence of fictional flights, resulting in the detection of borderline anomalous events that were mostly not reported by the baseline group.

Furthermore, five out of eight participants in the fictional flights group submitted reports of at least one of three ``close-call" anomalous events that eventually did not result in a fault, while only one participant in the baseline group did so. On the one hand, this indicates that fictional flights helped maintain the conflict prediction and situation assessment capabilities of participants. On the other hand, results also show that the presence of fictional flights is not perceived as being useful from an information flow perspective. This shows that an effect of gamification is that more information needs to be processed that is not directly useful for the actual goal. Here, participants were receiving and processing information about fictional flights that was not useful in supervising real flights. Thus, fictional flights may mitigate the effects of boredom, but also become a distracting element, as mental capacity is directed towards solving fictional conflicts.

The results from the SART questionnaire show that participants in the fictional aircraft group experienced a better concentration level (associated with mental workload), as increasing mental workload was one of the mechanisms through which gamification was expected to be beneficial in mitigating the effects of boredom. However, the trend observed in the information usefulness answers shows one of the dangers of overlaying fictional flights on top of real flights: the screen itself contains more information, but this information does not contribute to understanding the (real) situation. This is backed by some participants in the fictional flights group reporting that they found fictional flights distracting. Thus, for the highly automated ATC environments with nominal traffic density simulated in the experiment at hand, a beneficial increase in the level of situation awareness was not observed. Thus, hypothesis **HP-4** cannot be accepted.

Overall, the results obtained in this paper show that the effect of the use of fictional aircraft in a highly automated ATC environment is dependent on the personal preference of the users. For some experiment participants, it achieved an improvement in supervisory control performance in monotonous task situations, while others did not find its use beneficial. Gamification can be perceived both positively and negatively, depending on a multitude of factors, including background and personal strategy, and will thus inherently be experienced differently by potential users of a gamified tool.

Furthermore, several participants indicated that the presence of fictional aircraft was a distracting element, indicating that the use of such a tool in high-density, automated environments can be detrimental. This could be mitigated by allowing operators to toggle and configure the tool in function of need. Thus, in situations where fictional aircraft would be more distracting than beneficial, their number could be reduced, or they could be disabled altogether.

Lastly, the potential improvement in performance and ability to concentrate in low workload situation that some ATCOs might experience is worth pursuing in future research in this domain. In low traffic density situations (e.g., nighttime or extraordinary events such as the COVID-19 pandemic), skill degradation is a prevalent issue that is acknowledged by a majority of ATCOs (Kenny and Li, 2022). In such situations, a gamified tool such as the one investigated in this work could be used to increase the workload of ATCOs while maintaining their attention and engagement within the work environment of the main control task. As most participants indicated that they treated fictional aircraft similarly to real aircraft, this concept could be used to simulate nominal-workload situations to mitigate both skill degradation and the negative effects of a low-workload environment (e.g., boredom, low situation awareness). However, further design iterations of such a tool are required in order to ensure that it does not produce the negative effects experienced by some of the experiment participants (i.e., distraction and information clutter).

7 Conclusion

The work at hand sought to test the effects of the use of gamification, implemented as fictional flights, on ATCOs in a highly automated ATC environment. Sixteen students and staff members from the Faculty of Aerospace Engineering of TU Delft participated in an experiment. Participants had to detect anomalies in a simulated automated ATC system that issued commands to real flights, and had to take over manual control when a predetermined failure event occurred. A baseline group performed the supervisory control tasks without the presence of fictional flights on the screen, while a fictional flights group had to manually route them through the sector, avoiding conflicts with both types of other flights.

The subjective results obtained in this paper show that the use of fictional flights did not yield a significant improvement or deterioration in supervisory control performance in monotonous task situations. Gamification can be perceived both positively and negatively, depending on a multitude of factors, including the background and personal strategy. Thus, a tool that uses this strategy to enhance cognitive abilities could be implemented so that it can be toggled on and off. This would allow controllers to enable and disable the tool depending on the real traffic situation and personal preference. However, more research is needed to determine whether the inclusion of such tools in high traffic density situations is beneficial.

One of the main findings of the investigation at hand is given by the manner in which participants in the fictional aircraft group treated flights throughout the experiment. Most participants in this group indicated that they treated fictional aircraft as real aircraft unless their attention was required to supervise anomalous traffic situations. The introduction of gamified elements that are similar and relevant towards the main control task can help with simulating realistic situations and increase workload. While the results of the experiment do not support the applicability of such a tool within a high traffic density automated ATC environment, its use in low workload conditions might be beneficial. In situations that result in low traffic densities within a sector (e.g., nighttime or extraordinary events such as the COVID-19 pandemic), the voluntary deployment of fictional aircraft could maintain the situation awareness and engagement of controllers while also mitigating boredom and skill degradation.

Further research should be performed to better understand the potential benefits of using such tools and to discover ways in which the negative effects could be mitigated. Professional ATCOs should be used to obtain a more in-depth understanding of how cognitive skills are influenced by the proposed tool in a more realistic ATC environment. Furthermore, future experiments should collect more data (e.g., eye tracking) to improve measurements quantifying attention and vigilance.

Other game elements and strategies should also be explored (e.g., scores and achievements) in a wider range of situations, including normal ATC operation. However, this should be done while considering the ethical implications of modifying a safety-critical workflow that has evolved to a high standard of safety over decades. Investigations should be performed on whether the same vigilance improvement effects can be obtained in ways more compatible with current ATC work environments.

Contributor Statement

C.A. Badea: Conceptualisation, Data Curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualisation, Writing – Original Draft, Writing – Review & Editing

G. de Rooij: Conceptualisation, Formal analysis, Methodology, Software, Supervision, Resources, Validation, Visualisation, Writing – Review & Editing

C. Borst: Conceptualisation, Funding acquisition, Project administration, Software, Supervision, Resources, Writing – Review & Editing

M. Mulder: C. Borst: Conceptualisation, Funding acquisition, Project administration, Supervision, Resources, Writing – Review & Editing

Conflict Of Interest (COI)

There are no conflicts of interest.

References

Bainbridge, L. "Ironies of automation". In: *Automatica* 19.6 (1983), pp. 775–779. DOI: 10.1016/0005-1098(83)90046-8.

Bekier, M., Molesworth, B.R.C., and Williamson, A. "Tipping point: The narrow path between automation acceptance and rejection in air traffic management". In: *Safety Science* 50.2 (2012), pp. 259–265. DOI: 10.1016/j.ssci.2011.08.059.

Bekier, M. and Molesworth, B.R.C. "Altering user' acceptance of automation through prior automation exposure". In: *Ergonomics* 60.6 (2017), pp. 745–753. ISSN: 0014-0139. DOI: 10.1080/00140139.2016.1216610. (Visited on 2024-10-21).

Berberian, B., Ohneiser, O., De Crescenzio, F., Babiloni, F., Di Flumeri, G., and Hasselberg, A. "MINIMA Project: Detecting and Mitigating the Negative Impact of Automation". In: *Engineering Psychology and Cognitive Ergonomics: Performance, Emotion and Situation Awareness*. 2017, pages 87–105.

Borst, C., Flach, J.M., and Ellerbroek, J. "Beyond Ecological Interface Design: Lessons From Concerns and Misconceptions". In: *IEEE Transactions on Human-Machine Systems* 45.2 (2015). DOI: 10.1109/THMS.2014.2364984.

Borst, C., Visser, R.M., Paassen, M.M. van, and Mulder, M. "Exploring Short-Term Training Effects of Ecological Interfaces: A Case Study in Air Traffic Control". In: *IEEE Transactions on Human-Machine Systems* 49.6 (Dec. 2019), pp. 623–632. DOI: 10.1109/THMS.2019.2919742.

Cutler, V. and Paddock, S. "Use of threat image projection (TIP) to enhance security performance". In: *43rd Annual 2009 International Carnahan Conference on Security Technolog*y. 2009, pp. 46–51. DOI: 10.1109/CCST.2009.5335565.

EUROCONTROL. Situation Awareness Rating Technique (SART) — HP repository. 2012. url: http://web.archive.org/web/20200527113214/https://ext.eurocontrol.int/ehp/?q=node%2F1608 (visited on 2020-04-24).

Foroughi, C.K., Sibley, C., Brown, N.L., Rovira, E., Pak, R., and Coyne, J.T. "Detecting automation failures in a simulated supervisory control environment". In: *Ergonomics* 62.9 (2019), pp. 1150–1161. DOI: 10.1080/00140139.2019.1629639.

Hancock, P.A. "In Search of Vigilance The Problem of Iatrogenically Created Psychological Phenomena". In: *The American psychologist* 68 (2012). DOI: 10.1037/a0030214.

Hendrickx, P. and Tisza, A. B. EUROCONTROL Maastricht Upper Area Control Centre OPS & Automation Strategy. 2019. url: https://skybrary.aero/bookshelf/books/5341.pdf.

Hofer, F. and Schwaninger, A. "Using threat image projection data for assessing individual screener performance". In: *WIT Transactions on the Built Environment* 82 (Jan. 2005), pp. 417–426. DOI: 10.2495/SAFE050411.

Hurter, C., Degas, A., Guibert, A., Durand, N., Ferreira, A., Cavagnetto, N., Islam, M. R., Barua, S., Ahmed, M.U., Begum, S., Bonelli, S., Cartocci, G., Flumeri, G.D., Borghini, G., Babiloni, F., and Aric´o, P.. "Usage of more transparent and explainable conflict resolution algorithm: air traffic controller feedback". In: *Transportation Research Procedia. 34th Conference of the European Association for Aviation Psychology* 66 (2022), pp. 270–278. ISSN: 2352-1465. DOI: 10.1016/j.trpro.2022.12.027.

Kenny, Shane and Li, Wen-Chin. "The Effect of the COVID-19 Pandemic on Air Traffic Controllers' Performance: Addressing Skill Fade Concerns in Preparation for the Return to Normal Operations". In: *Engineering Psychology and Cognitive Ergonomics*. Vol. 13307. Cham: Springer International Publishing, 2022, pp. 300–313. ISBN: 978-3-031-06085-4 978-3-031-06086-1. DOI: 10.1007/978-3-031-06086-1_23.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovi, Z., Jaskolski, M., and Baker, D. "Crystal structure of a monomeric retroviral protease solved by protein folding game players". In: *Nature structural & molecular biology* 18.10 (Sept. 18, 2011), pp. 1175–1177. DOI: 10.1038/nsmb.2119.

Kim, T. and Werbach, K. "More than just a game: ethical issues in gamification". In: *Ethics and Information Technology* 18.2 (June 2016), pp. 157–173. DOI: 10.1007/s10676-016-9401-5.

Langford, K., Kille, T., Lee, S., Zhang, Y., and Bates, P.R. ""In automation we trust" - Australian air traffic controller perspectives of increasing automation in air traffic management". In: *Transport Policy* 125 (2022), pp. 352–362. ISSN: 0967-070X. DOI: 10.1016/j.tranpol.2022.07.001.

Loft, S., Bhaskara, A., Lock, B.A., Skinner, M., Brooks, J., Li, R., and Bell, J.. "The Impact of Transparency and Decision Risk on Human–Automation Teaming Outcomes". en. In: *Human Factors* 65.5 (2023), pp. 846–861. ISSN: 0018-7208. DOI: 10.1177/00187208211033445.

Mackworth, N. "The Breakdown of Vigilance during Prolonged Visual Search". In: *Quarterly Journal of Experimental Psychology* (1948). DOI: 10.1080/17470214808416738.

Mercer, J., Gomez, A., Gabets, C., Bienert, N., Edwards, T., Martin, L., Gujral, V., and Homola, J. "Impact of Automation Support on the Conflict Resolution Task in a Human-in-the-Loop Air Traffic Control Simulation". In: *13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (2016). DOI: 10.1016/j.ifacol.2016.10.458.

Meuter, R. and Lacherez, P. "When and Why Threats Go Undetected: Impacts of Event Rate and Shift Length on Threat Detection Accuracy During Airport Baggage Screening". In: *Human Factors* 58.2 (Mar. 1, 2016), pp. 218–228. DOI: 10.1177/0018720815616306.

Munyung, K., Borst, C., and Mulder, M. "Situation Awareness Prompts: Bridging the Gap between Supervisory and Manual Air Traffic Control". In: *15th IFAC Symposium on Analysis Design and Evaluation of Human Machine Systems*. 2022.

Nieto, S. and Javier, F. "The long journey toward a higher level of automation in ATM as safety critical, sociotechnical and multi-Agent system". In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering* 230.9 (July 1, 2016), pp. 1533–1547. DOI: 10.1177/0954410015596763.

Noskievic, T. and Kraus, J. “Air Traffic Control Tools Assessment”. In: *MAD - Magazine of Aviation Development* 5.2 (2017), pp. 6–10. DOI: 10.14311/MAD.2017.02.01.

Ohneiser, O., De Crescenzio, F., Di Flumeri, G., Kraemer, J., Berberian, B., Bagassi, S., Sciaraffa, N., Aric`o, P., Borghini, G., and Babiloni, F. “Experimental Simulation Set-Up for Validating Out-Of-The-Loop Mitigation when Monitoring High Levels of Automation in Air Traffic Control”. In: *International Journal of Aerospace and Mechanical Engineering* 12.4 (2018), p. 12.

Paja, E., Poggianella, M., Aydemir, F., and Giorgini, P. PACAS: *A Gamified Platform for Participatory Change Management in ATM Systems. Proceedings of International Conference on Requirements Engineering: Foundation for Software Quality* (REFSQ 2018), Utrecht, The Netherlands, 2018.

Parasuraman, R., Molloy, R., and Singh, I.L. “Performance Consequences of Automation-Induced ‘Complacency’”. In: *The International Journal of Aviation Psychology* 3.1 (1993), pp. 1–23. DOI: 10.1207/s15327108ijap0301_1.

Parasuraman, R., Sheridan, T.B., and Wickens, C.D. “A model for types and levels of human interaction with automation”. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 30.3* (2000). DOI: 10.1109/3468.844354.

Pop, V., Stearman, E., Kazi, S., and Durso, F. “Using Engagement to Negate Vigilance Decrements in the NextGen Environment”. In: *Int. J. Hum. Comput. Interaction 28* (Feb. 1, 2012), pp. 99–106. DOI: 10.1080/10447318.2012.634759.

Rogosic, T., Juricic, B., Antulov-Fantulin, B., and Radisic, T. “Review of Automation Problems in Air Traffic Control”. In: *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. 2021, pp. 1636–1641. DOI: 10.23919/MIPRO52101.2021 .9597063.

Rooij, G. de, Tisza, A.B., and Borst, C. “Flight-Based Control Allocation: Towards Human–Autonomy Teaming in Air Traffic Control”. In: *Aerospace* 11 (2024), pp. 919. DOI: 10.3390/aerospace11110919

Sargent, R., Walters, B., and Wickens, C.D.. “Meta-analysis Qualifying and Quantifying the Benefits of Automation Transparency to Enhance Models of Human Performance”. en. In: *Human-Computer Interaction*. Ed. by Masaaki Kurosu and Ayako Hashizume. Cham: Springer Nature Switzerland, 2023, pp. 243–261. ISBN: 978-3-031-35596-7. DOI: 10.1007/978-3-031-35596-7_16.

Schwaninger, A. “Threat Image Projection: Enhancing Performance?” In: *Aviation Security International* (Dec. 2006), pp. 36–41.

SESAR Joint Undertaking. *European ATM Master Plan: Digitalising Europe’s Aviation Infrastructure*. Publications Office of the European Union, 2019. ISBN: 978-92-9216-110-1.

Smith, P.J. and Baumann, E.. “Human-Automation Teaming: Unintended Consequences of Automation on User Performance”. In: 2020 *AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*. 2020, pp. 1–9. DOI: 10.1109/DASC50938.2020.9256418.

Strauch, B. “Ironies of Automation: Still Unresolved After All These Years”. In: *IEEE Transactions on Human-Machine Systems* 48.5 (2018). DOI: 10.1109/THMS.2017.2732506.

Svensson, A, Lundberg, J., Forsell, C., and Ronnberg, N.. “Automation, teamwork, and the feared loss of safety: Air traffic controllers’ experiences and expectations on current and future ATM systems”. In: *Proceedings of the 32nd European Conference on Cognitive Ergonomics*. ECCE ’21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–8. ISBN: 978-1-4503-8757-6. DOI: 10 . 1145/3452853.3452855.

Volz, K.M. and Dorneich, M.C. “Evaluation of Cognitive Skill Degradation in Flight Planning”. In: *Journal of Cognitive Engineering and Decision Making* 14.4 (2020), pp. 263–287. ISSN: 1555-3434. DOI: 10.1177/1555343420962897.

Wickens, C.D., Gordon, S., and Liu, Y. *An Introduction to Human Factors Engineering*. 2nd. Pearson Education, 2004. ISBN: 978-0131837362.

Zichermann, G. and Cunningham, C. *Gamification by Design. First Edition.* Sebastopol, California, USA: O'Reilly Media, 2011. ISBN: 978-1-4493-9767-8.