Gamification: Improving Supervisory Control Performance in Highly Automated Air Traffic Control

Master of Science Thesis

C. A. Badea

Technische Universiteit Delft

TU Delft

Gamification: Improving Supervisory Control Performance in Highly Automated Air Traffic Control

Master of Science Thesis

by

C. A. Badea

Student number:	4439767	
Project duration:	December, 2019 – January, 2021	
Thesis committee:	Dr. ir. C. Borst,	TU Delft, daily supervisor
	Prof. M. Mulder,	TU Delft, supervisor
	Dr. ir. B.F. Santos,	TU Delft, external examiner
	Ir. G. de Rooij,	TU Delft, additional member

TU Delft

List of Figures

List of Tables

Contents

I

Scientific Paper

Using Gamification For Improving Supervisory Control Performance in Highly Automated Air Traffic Control

Calin Andrei Badea (MSc Student)
Supervisors: Dr. Ir. Clark Borst, Prof. Dr. Ir. Max Mulder, and Gijs de Rooij, MSc
Control and Simulation Department of the Faculty of Aerospace Engineering, TU Delft

Abstract—**The use of automated conflict detection and resolution tools for 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 domains such as luggage screening, where dangerous items are rarely found, by projecting fictional threats on top of luggage scans. This paper presents and experimentally tests a proposed implementation of gamification within an air traffic control work environment. Fictional aircraft are superimposed among automatically controlled real traffic, thus creating fictional conflicts that need resolving. System supervisors are given the task of supervising the behaviour of a fully automated conflict detection and resolution system while routing fictional aircraft safely and efficiently through the sector, avoiding conflicts with other aircraft (both real and fictional). Automation anomalies are simulated during the experiment, as well as an automation failure event after which the system supervisor needs to take over manual control of traffic. Experiment results show that the presence of fictional aircraft increased reported concentration levels among participants, as well as improved supervisory control performance. No significant differences have been recorded in manual control performance, however some participants reported that fictional aircraft were distracting. Thus, while the use of fictional aircraft increases engagement, it might negatively affect other cognitive functions. Further research is recommended to investigate the use of gamification within the safety-critical ATC work environment.**

1. Introduction

The aviation industry has always striven for maximising the efficiency and safety of commercial operations, driven by the introduction of advanced technologies both in the cockpit and on ground. The air traffic control domain is predicted to undergo fundamental modernisation in the next 20 years, as the push for increased automation is gaining traction. Controller aiding tools are currently being used, such as trajectory prediction and short term collision alerting [1], however the decision-making process still rests upon the air traffic controller.

The SESAR Air Traffic Control (ATC) Master Plan of 2019 [2] mentions that a high level of automation in air traffic control will be reached by the year 2040, facilitated by the development and widespread use of ADS-B technology. The use of automatic conflict detection and resolution systems implies that air traffic controllers will undertake the role of supervisors, intervening in case of the occurrence of exceptional events, such as automation failures. This will have major implications on the work environment, as the nature of the control task fundamentally changes.

The Ironies of Automation, described by Brainbridge [3], describe potential issues that might arise when automating tasks previously performed by humans. Although much research has been conducted in the field of automation since, the issues have not been resolved [4]. Several negative effects of the practice of highly monotonous supervisory tasks by humans are described by Parasuaraman et al. [5]. 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, and thus are less aware of the decisions automation makes and their consequences. Another issue mentioned by Parasuraman stems from the high reliability of automated systems in safety-critical applications, which could produce complacency in supervisors and thus a higher probability of missing potential failures.

Several solutions to the cognitive issues posed by automation have been proposed and researched. Mercer et al. [6] showed that simply involving controllers in the conflict detection task (where conflict resolution is fully automated) improves situation awareness. Pop et al. [7] showed that producing and maintaining engagement benefits failure detection in air traffic control supervisors, and is a key element towards improved cognitive abilities. Other potential methods through which cognitive performance could be improved is by aiding controllers in understanding the actions of automation and provide more transparency. Borst et al. [8] propose the use of ecological interface design to help system supervisors detect faults in automation by increasing transparency and information flow within the human-machine system.

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. One potential solution to this could be the use of gamification (the use of game elements in non-gaming contexts [9]). Threat image projection (TIP), identified as a form of gamification, is a technique used in airport luggage screening (a highly monotonous task), and implies the superposition of fictional threats (such as firearms) onto luggage x-ray scans. Thus, the rate at which operators are exposed to threats is increased, which mitigates the effects of boredom.

This paper proposes the adaptation and implementation of TIP within a highly automated ATC environment through the use of fictional aircraft introduced among real traffic, the latter being controlled by a fully automatic conflict detection and resolution system. Fictional aircraft would require manual control and the development and maintenance of a mental model of the aircraft in the sector, thus achieve higher engagement while mitigating the effect of boredom.

The increase in engagement would have a positive effect on supervisory control performance, and seeks to improve the transition from supervisory to manual control in case of automation failure by maintaining the operator in the control loop. The proposed concept was tested experimentally using 16 participants.

The article is structured as follows: Section 2 analyses the current trend in automation, and the implications of gamification. Section 3 presents the design considerations and the proposed implementation of fictional aircraft within an ATC environment. The experiment method and the obtained results are presented in Sections 4 and 5. Lastly, Sections 6 and 7 present the conclusions that arise from the obtained data, as well as recommendations for future research.

2. BACKGROUND

Automation in ATC

As technological developments rapidly advance, the push for the introduction of automation in air traffic control is gaining traction. According to Nieto [10], high levels of automation in ATC will result in an increase in safety and airspace capacity. However, this entails that the decision-making process must be shifted from human operators to computers. Operators will assume the role of system supervisors [11], a task that Foroughi et al. have experimentally determined to be unsuitable for humans [12].

The main source of human incompatibility with supervisory tasks stems from the 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 [13]. Situation awareness and vigilance are therefore negatively impacted, which leads to two types of human errors, as documented by Berberian et al. within the MINIMA project [14], 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. [14] mention the current proposed solutions for maintaining air traffic controllers 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, a fourth solution that has the potential to facilitate the mitigation of out-of-the-loop issues is the concept of gamification.

Concept of gamification

Gamification is a technique through which game elements are used in non-gaming contexts with the purpose of obtaining the cognitive benefits that humans experience in such contexts. Elements such as leader-boards, scores and achievements can increase productivity and motivation by creating a sense of progression and reward [9]. Gamification is currently being used in a wide range of domains, such as management techniques (PACAS [15] project, which proposes a change in the management style of ATM) or research (the use of mobile games for the simulation of HIV virus behaviours [16]). The latter makes use of a gamification technique through which fictional goals (completing puzzles), which are more appealing and motivating to the general public, are indirectly used to fulfil real goals (HIV virus simulation).

One successful implementation of gamification within the aviation industry is the use of threat image projection (TIP) 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 [17] proved that long shifts lead to an increase in threat detection errors. To mitigate this, some airports have implemented threat image projection, described by Schwaninger [18] as the projection of fictional images of threats on the 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 imposed over regular images of luggage. This superposition is also varied in its complexity and level of obstructiveness, achieving a wide range of view difficulty and complexity. According to Hofer and Schwaninger [19], the increase in threat rates lead to better detection performance if the fictional object database is diverse, thus providing a vast variety of exceptional situations.

Compared to other types of gamification, threat image projection is implemented in a safety-critical environment, facilitating the porting of the concept to air traffic control. The gamification elements and concepts used within TIP have the potential to increase the engagement of supervisors 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 better understanding of the situation as well as better automation fault detection rates.

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 air traffic control context, two ethical considerations must be taken into account when designing an implementation. Firstly, according to Kim and Werbach [20], gamification can create a manipulative environment in which participants are forced to partake into a game without their consent. This is apparent when considering threat image projection, as screening operators do not have a choice when it comes to participation. Furthermore, gamification leads to the creation of fictional goals, which can be manipulative if participants are not aware of the nature of these goals.

The second ethical issue with gamification is that it can produce unnecessary stress to participants as the stakes are artificially raised [20]. In threat image projection, the operators do not have prior knowledge about the nature of a threat, and are only told whether it was fictional or not after it is identified. 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.

3. Design and Implementation

The main feature of threat image projection is that fictional threats are introduced to the work space of a system supervisor. Within the air traffic control domain, the most prominent threat to the overall safety level of operations are aircraft conflicts. The proposed implementation of gamification revolves around the use of fictional aircraft superimposed onto real traffic to create virtual conflicts that require operator intervention, thus maintaining a higher level of engagement when compared to a purely supervisory task. Unlike real aircraft, which would be controlled by automation, fictional aircraft must be manually commanded by system supervisors.

Control task analysis

The intended effect of the use 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. [21] 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 is required to understand the current state of traffic and formulate a command. However, the use of fictional aircraft can potentially increase situation awareness, thus the first half of the decision making process will already be performed before the failure event if the operator is aware of the situation and consequences. This could lead to better performance when transitioning from supervisory to manual control, as shortcuts to points **B** or **C** are enabled.

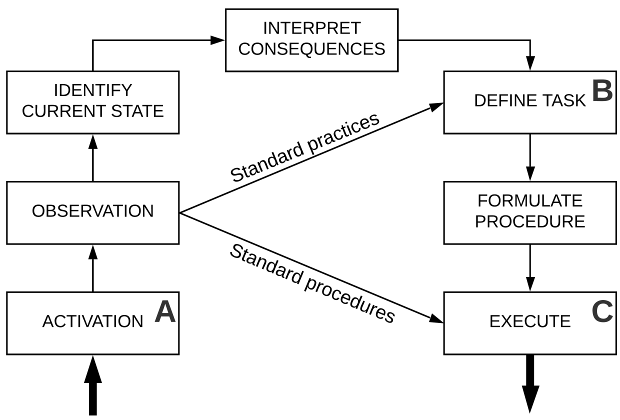


Figure 1: Simplified decision ladder diagram [21]. Point **A** is the standard entry point when transitioning from manual to supervisory control. If the human operator is kept within the control loop, shortcuts after activation to point **B** or point **C** are more accessible.

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 conflict detection and resolution system with the help of ATC tools (e.g., short term collision avoidance alerts) while also performing the task of manually controlling fictional aircraft. The latter enables the operator to perform a secondary task within the work space that exercises manual control skills and requires building a metal model encompassing both real and fictional aircraft. The second type of control task that operators must perform in case of automation failure is the manual control of real aircraft, which does not differ from the current non-automated air traffic control task. Fictional aircraft would not be a part of the control task in order to avoid increasing the workload over a manageable threshold for controllers.

Work space integration

The integration of fictional aircraft within the ATC work space was influenced by the previously presented ethical considerations of gamifying safety critical work environments. Firstly, fictional aircraft were implemented such that operators would be aware of their nature and be distinguishable from real aircraft, thus increasing transparency compared to TIP, where operators are not initially aware of the fictional nature of threats. This also facilitates the differential prioritisation of tasks. Both supervising real aircraft and manually controlling fictional aircraft share the same goal, routing aircraft through the sector safely and efficiently. Furthermore, in a potential implementation of gamification, operators would have the ability to activate or deactivate fictional aircraft in order to ensure that the use is voluntary. By considering the display design principles of discriminability and redundancy described by Wickens [22], the design presented in Figure 2 was created, in which fictional aircraft differ from real aircraft in both colour and shape (icon). The colour blue was chosen due to its high contrast with the dark backgrounds usually used in ATC displays.

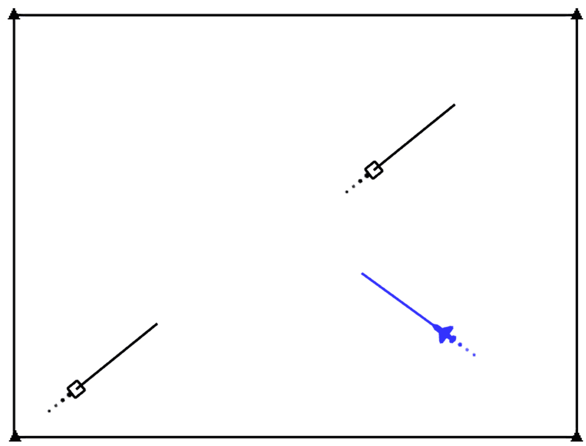


Figure 2: Simplified visualisation of fictional aircraft introduced on the ATC radar screen.

By analysing the design considerations of threat image projection, two main factors that affect the influence of fictional aircraft 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 aircraft (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 aircraft will travel, there are three types of traffic patterns relative to these routes: (1) fictional aircraft are introduced such that they cross the main traffic flow, (2) fictional aircraft are introduced among the aircraft 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 aircraft 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 aircraft demands higher workload, thus making the supervisory task more difficult to perform. This is undesirable, as supervising air traffic is the highest priority. The hypothesised effect of combinations of these two design considerations are presented in Table 1. Three human factors were taken into account: situation awareness (SA), workload (WL) and control performance (PF).

Table 1: Expected effect of fictional aircraft pattern and density on situation awareness, workload, and control performance compared to the unaided supervisory control case.

Density	Traffic pattern		
	Crossing	Among	Combined
Low	SA ↑	SA =	SA ⇈
	WL ↑	WL ↑	WL ↑
	PF ↑	PF =	PF ⇈
High	SA ⇊	SA =	SA ↓
	WL ⇈	WL ↑	WL ↑
	PF ⇊	PF =	PF ↓

In Table 1, up and down arrows represent an expected increase or decrease in the corresponding factor (double arrows represent a large increase or decrease), the equal sign means the factor remains approximately unchanged, and the colours portray if the change is beneficial or detrimental to the supervisory control performance of operators.

It is expected that the presence of fictional aircraft will always lead to higher workload when compared to an unaided supervisory control situation. However, if all fictional aircraft 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 aircraft around them. On the other hand, if all fictional aircraft are crossing the main traffic flow, more conflicts will arise and they will require much more maintenance and attention, thus leading to increased workload.

From the point of view of situation awareness, a combined fictional aircraft traffic pattern will produce more evenly distributed traffic in the sector, thus aiding the operator in maintaining an overview of all parts of the sector. Lastly, if the number of fictional aircraft in the sector is high compared to the number of real aircraft, more resources will have to be directed to the manual control task, which will distract air traffic controllers and lower situation awareness. Therefore, the combination with the greatest potential for improving performance is hypothesised to be a low number of fictional aircraft that are introduced both within and crossing the main traffic flows in the sector.

Depending on the way in which automation is implemented within the air traffic control system, supervisors would be able to take over manual control either through their own decision or in cases when their intervention would be required (for example, an automation fault detection system). In these cases, is it undesirable to have fictional aircraft present on the screen, as they would provide little benefit or even be distracting. Thus, they would automatically disappear from the screen to allow the air traffic controller to allocate their full attention towards real aircraft.

4. Method and Experimental Setup

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

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

Apparatus and software

The experiment was conducted using a modified version of SectorX, a TU Delft in-house developed Java based air traffic control simulator. The simulator ran on a single computer setup, as shown in Figure 3, including a traditional LCD display (76 cm, 1920 × 1200 pixels), a mouse, and a keyboard. The environment was set up such that potential distractions were minimised (window view was obstructed, distracting elements such as clocks were removed from the desktop environment).

SectorX was modified to include fictional aircraft, a supervisory control mode and a manual control mode. During the supervisory phase of a scenario, automation was enabled and handled all real aircraft, while only fictional aircraft could be manually controlled if present. The manual control phase began when the scenario automation failure time is reached. Fictional aircraft disappeared from the screen, and real aircraft manual control was enabled. The simulator was therefore augmented to accommodate anomaly reporting during the supervisory control phase. While automation was active, unusual behaviour of real aircraft could be reported by clicking the offending aircraft and typing a report, as seen in

Figure 3: The hardware setup used for the experiment.

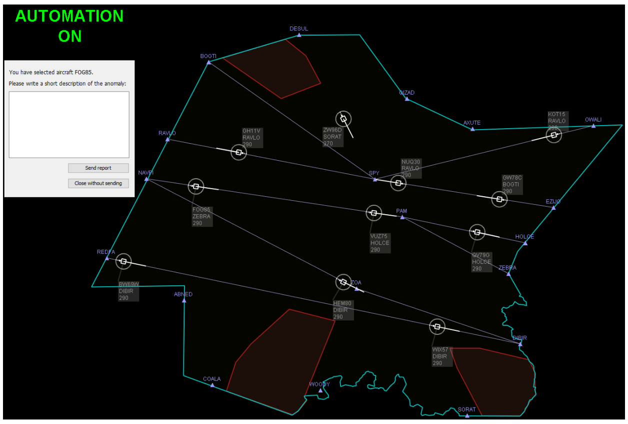


Figure 4: Screenshot of SectorX (t = 0 s) while automation is active. “Automation on” text is displayed on screen. If an aircraft is clicked, the anomaly report window appears.

Figure 4. Furthermore, when automation failed, a notification appeared on screen, which needed to be acknowledged and dismissed by clicking on it, after which the ability to manually control aircraft was enabled, as shown in Figure 5.

Participants

Sixteen participants volunteered to participate in the experiment, most part of the Control and Simulation department of the Faculty of Aerospace Engineering of TU Delft where air traffic management research is conducted. The participants had various backgrounds and experience levels with air traffic control: nine master students, five doctorate students (one of which undertook air traffic control training) and two lecturers. Among master students, most had previous air traffic control experience through the means of university courses as well as previous experience with the SectorX ATC simulation environment.

Independent variables

The experiment aimed to determine the effects of introducing fictional aircraft within a highly automated air traffic control environment. Thus, one independent between-participants variable was selected: the presence of fictional aircraft, limiting the complexity of the experiment and increasing the probability of capturing significant differences between groups. A between-participants experiment design was therefore used to ensure that all participants would experience the same level of real aircraft traffic and the same anomalies. The resulting experiment matrix is presented in Table 2.

Figure 5: Screenshot of SectorX immediately after failure event (t = 4000 s). The automation failure alert appears on screen and needs to be dismissed. Main routes are removed to encourage participants to route aircraft directly towards exit waypoints.

Table 2: One module of the experiment matrix. Participants are divided in two groups, therefore an even number of total participants was needed.

Participant No.	Scenario	Duration [s]
1	Fictional aircraft	4700
2	Baseline	4700
⋮	⋮	⋮

The participants were divided into two groups. Group A was given the task of manually controlling fictional aircraft while performing the supervisory control task. Group B, the control group, had to supervise the aircraft on the radar screen without the presence of fictional aircraft. The two groups were formed by considering the level of experience and knowledge in the air traffic control domain of the participants: Group A consists of five master students, two doctorate students and one lecturer, and Group B consists of four master students, three doctorate students and one lecturer.

Participant goals and tasks

Participants were instructed to perform the following tasks:

- **Primary supervisory control phase task:** supervise traffic controlled by automation in sector, 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 task:** route fictional aircraft, if present, safely and efficiently towards their exit waypoints. Avoid conflicts with other real and fictional aircraft. Automation does not account for the pres-

ence of fictional aircraft, thus compensate and command fictional aircraft in case a conflict arises as a result of an automatic aircraft manoeuvre.

- **Primary manual control phase task:** after the automation failure, dismiss the notification as soon as possible and proceed with routing aircraft towards their exit waypoints safely and efficiently until the end of the traffic scenario. Automation will not be re-enabled.

Air traffic scenarios

The Delta sector of the Maastricht Upper Area Control Centre (MUAC) was selected for the experiment due to the familiarity of the participants with the Dutch airspace and data availability for this sector. ADS-B data from the year 2018 was analysed and used to develop realistic traffic patterns. The most popular flight routes can be observed in Figure 6, in which the heat map of aircraft positions throughout the day of the 25th of July 2018 is plotted. It should be noted that the heat map emphasises intersection points more than the rest of the air routes.

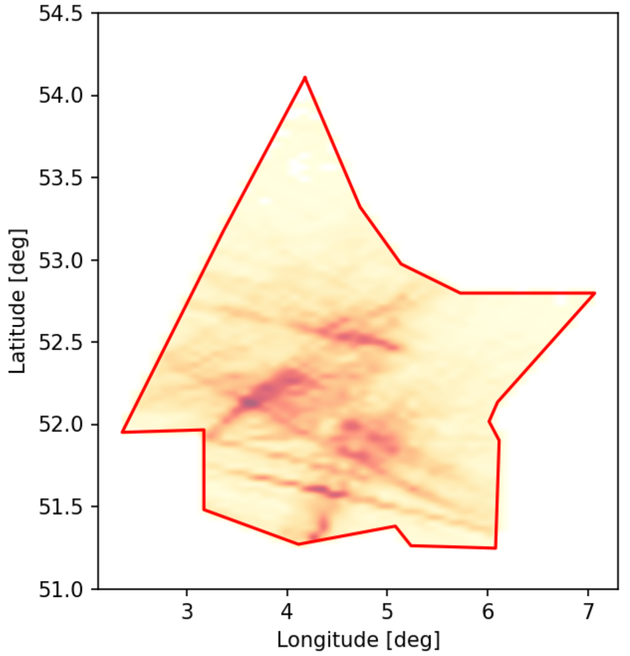


Figure 6: ADS-B data heat map for the 25th of July 2018 collected by receivers at TU Delft, altitude over FL245. Less data is present in the East and North of the sector due to receiver range limitations.

The ADS-B data was used to create realistic traffic patterns within the Delta sector: the main traffic flow would cross the sector in both directions of the East-West plane, while non-regular aircraft would cross the sector and main traffic flow in the North-South plane. This produces regular disruptions that require the attention and supervision of an operator.

Restricted airspace was also added to the sector as an extra factor to be considered when supervising. Three such areas were placed at the edges of the sector as to not cause disruptions to the main flow, as well as to test the situation awareness of participants with events occurring away from the centre of the screen. These are based on actual restricted airspace locations that can be found within the sector, mainly military airspace and special use airspace. Therefore, the final sector configuration is presented in Figure 7. The northern part of the sector was cropped to allow a better fit on a screen, as well as due to the lack of ADS-B data available for that section of the sector. Exit and entry waypoints were distributed among the boundary of the sector, as well as three inner waypoints coinciding with the high density areas presented in Figure 7.

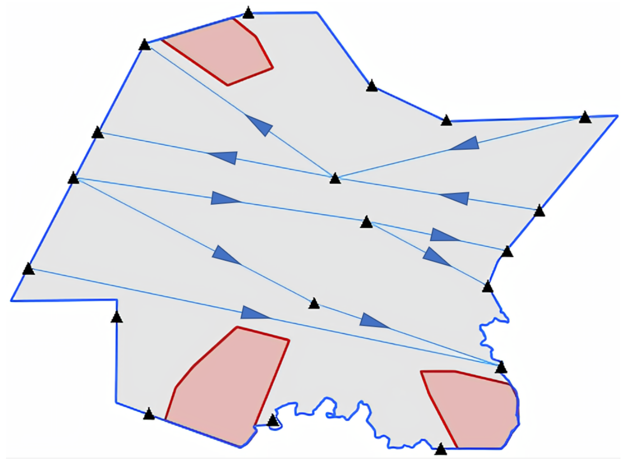


Figure 7: Modified Delta sector used for the experiment traffic scenario.

ADS-B data was analysed to gain an indication of the average aircraft per air traffic controller in the Delta sector in order to produce scenarios with realistic traffic density. According to EUROCONTROL [23], the Delta sector is divided into three layers: a low sector (FL245 - FL335), a middle sector (FL335 - FL365) and a high sector (above FL365). By analysing the data, an average of 45 aircraft per altitude layer per hour was calculated. The indicated air speed of aircraft following the main routes described in Figure 7 was set at 290 kts (444 kts ground speed at FL290). Thus, the sector is traversed within 15 minutes on average. Therefore, the baseline scenario was designed to have a traffic density of 45 aircraft per hour, achieved by maintaining approximately 11 aircraft within the sector at all times.

The length of the experiment traffic scenarios was decided upon based on the attention decrement phenomenon described by Mackworth [24], showing through an experiment that the greatest decrement in attention while performing a monotonous supervisory task occurs in the first hour. Hancock [25] expands on the concept and links the attention decrement to the nature of the task. Based on these considerations, the length of supervisory control part of the scenario was set at 4,000 seconds (66.7 minutes) and the length of the manual control part at 700 seconds (11.7 minutes).

Fictional aircraft scenarios

The fictional aircraft scenarios were developed based on the design considerations and the hypothesised ideal density. These scenarios were built by adding fictional aircraft to the baseline scenarios. Thus, between the two groups, the real aircraft would be the same. Through several iterations and preliminary test participants, it was decided that fictional aircraft would represent 20% of the real aircraft on screen, therefore approximately 2 fictional aircraft would be in the sector at all times.

Automation, anomalies and failure

The SectorX simulator is capable of recording and playing back commands of a manually controlled run. This was used to simulate an automated conflict detection and resolution algorithm by recording commands for the manual control part of the experiment scenario. The automated commands were given such that real aircraft would maintain a separation of 7 nautical miles. Aircraft would also mostly follow the main routes presented in Figure 7 and manoeuvre at the three interior waypoints when possible to increase automation transparency and predictability. Furthermore, the text on the labels of real aircraft changes colour while an automated command is executed, as shown in Figure 8. However, although desirable, participants were not given insight into the command content in order to avoid interference with the measurement of the effect of fictional aircraft on supervisory performance. Furthermore, the presence of fictional aircraft had no impact on the commands given to automation, thus conflicts could arise between fictional and real aircraft that need human intervention. Separation circles with a radius of 2.5 nautical miles, history dots and one minute look-ahead velocity vectors were also added to aircraft icons to facilitate the supervision task.

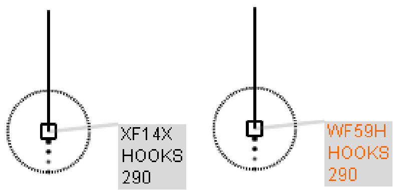


Figure 8: Standard aircraft label (left) and manoeuvring aircraft label (right). Label contains the following information: aircraft ID (first line), exit waypoint (second line) and indicated airspeed in knots (third line).

The anomalous events that were placed throughout the supervisory phase of the scenario were of three types:

1. **Loss of separation:** two aircraft breach the minimum lateral separation requirement of 5 NM
2. **Restricted airspace separation violation:** aircraft get close (less than 2.5 NM) or breach restricted airspace areas
3. **Wrong exit waypoint:** aircraft exits the sector through a different waypoint than assigned.

These three phenomena were selected due to the objectivity with which they could be spotted, and placed at various time intervals throughout the scenario. In total, seven anomalous events occur during the experiment, presented in Table 3. Two anomalies occur at the same time in different regions of the screen (at 3,015 seconds) to provide insight on the occurrence of attention tunnelling (i.e., attention is disproportionately drawn towards one part of the screen).

Table 3: Anomalous events that occur during the experiment traffic scenario.

ID	Time [s]	Anomaly Type
1	797	Restricted airspace violation
2	1,603	Wrong exit waypoint
3	2,033	Loss of separation
4	2,520	Restricted airspace violation
5	3,015	Restricted airspace violation
6		Wrong exit waypoint
7	3,870	Restricted airspace violation

Lastly, an automation failure event occurred at the 4,000 second mark of the scenario. This was implemented to test the performance of operators that have to transition from supervisory to manual control. It was designed on the premise that future implementations of automated conflict detection and resolution systems will have a very high degree of integrity, and will therefore be able to diagnose an internal issue and give manual control to the system supervisor. Thus, when the failure event occurred in the experiment scenario, an alert message was shown on the screen. The operator dismissed the alert by clicking on it and began manually controlling the aircraft in the sector.

When the automation failure event occurred, all fictional aircraft, if present, were deleted from the screen. This was done to let the participants dedicate all their resources towards the manual control of real aircraft, while also levelling the conditions in which manual control performance is measured. A visual summary of the scenario is presented in Figure 9, where "Total a/c" was the traffic level (real + fictional) seen by the fictional aircraft group, and "Real a/c" was the traffic level experienced by the baseline group.

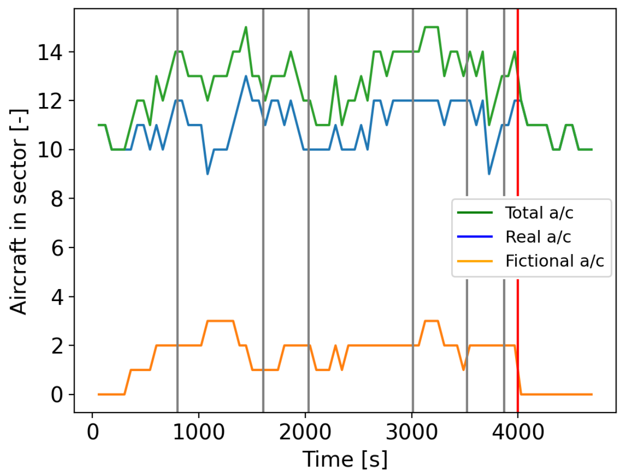


Figure 9: Summary of fictional aircraft and baseline scenarios. Gray vertical lines represent the locations of anomalies throughout the scenario, while the red vertical line represents the failure event.

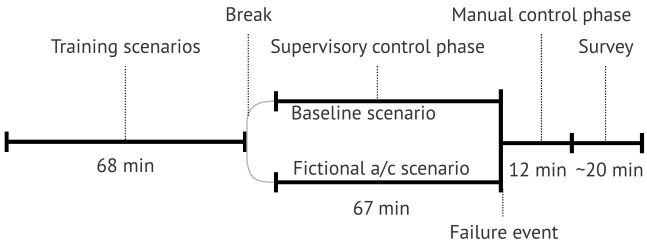


Figure 10: Experiment timeline for each participant. All participants undertook the same training, and the same setup for the manual control phase. The supervisory control phase differed depending on experimental group.

Procedure

Several days before an experiment session, the participants were sent a briefing manual containing information about the experiment setup, the simulator, as well as instructions on the experimental task. All participants regardless of group were given the same briefing and the same training procedure. Seven training scenarios were developed for this purpose. The first two introduced the basic controls and functionality of the simulator. Scenarios three to five consisted of manual control tasks with progressively increasing difficultly, aiming to familiarise participants with the simulation environment, the traffic density, and the sector that was used during the experiment. Scenarios six and seven introduced participants to fictional aircraft and automated conflict resolution for real traffic, as well as the anomaly report and failure alert features of the simulator. At the end of the last scenario, participants were shown a traffic situation in which all three types of anomalies were demonstrated. However, participants were encouraged to report all events that they deemed anomalous, regardless if the anomalies matched the types mentioned in the briefing.

The training process lasted for one hour, ensuring that all participants, regardless of their experience level in air traffic control, were familiar with the simulation environment as well as had enough experience to control traffic in an efficient and safe manner. Furthermore, all participants experienced the concept of fictional aircraft, leading to more insightful feedback in the survey conducted after the experiment.

After training, participants were briefed on which group they are part of (baseline or fictional aircraft group) and what their mission is during the experiment traffic scenario. While the participants were performing this scenario, interaction with the researcher was kept at a minimum. After the experiment traffic scenario, which lasted for approximately 1.3 hours, participants were asked to fill in a questionnaire. The whole experiment procedure lasted for approximately three hours for each participant, and is presented in Figure 10.

Control variables

Due to the novelty of the concept, as well as the expected high variability in air traffic control characteristics and experience among participants, a large number of variables were controlled during the experiment:

- *Real aircraft traffic:* All participants experienced the same real aircraft traffic during both the supervisory and manual control phases of the scenarios.
- *Air sector structure:* Both traffic scenarios used the same sector structure, including waypoint locations and names, main routes, and restricted airspace locations.
- *Degrees of freedom:* In order to increase the comparability of data between participants, the degree of freedom of aircraft was limited to heading only. Thus, only heading commands could be issued, and all aircraft flew at FL290. The indicated air speeds of aircraft varied between 250 kts and 310 kts, but could not be changed by participants or automation. Along the main routes, the indicated air speed of all aircraft (real and fictional) was 290 kts. Along crossing routes, the indicated air speed of aircraft varied.
- *Anomalous events:* Automation anomalies were exactly the same across the fictional aircraft and baseline scenarios.
- *Aircraft type:* All aircraft were of the same type, and thus had the same physical properties and performance.
- *Radar update rate:* The radar update frequency was set at once every 5 seconds (0.2 Hz) to match the previous experience of participants with SectorX.
- *Automation commands:* As both scenarios contained the same real aircraft, the scripted automation was the same for all participants.
- *Traffic situation after failure:* As the evolution of aircraft was the same regardless of scenario type, the traffic situation immediately after the failure event was the same for all participants, thus participants were faced with the same manual control challenge.
- *Training and briefing:* All participants regardless of group received the same briefing information and performed the same training scenarios, including scenarios that contained fictional aircraft.
- *Experiment environment:* All participants performed the experiment in the ATM Lab of the Faculty of Aerospace Engineering of TU Delft. The screen was oriented such that no distracting elements were in view, and the SectorX simulator was run in full screen mode to hide the task bar and clock of the operating system.
- *ATC simulation environment:* Features that would normally be present in an ATC simulation environment (e.g., short term conflict alert, conflict prediction tools) were deactivated for the experiment to eliminate the influence of confounding factors on the results.

Dependent measures

The following variables were measured:

- *Anomaly reports:* Both the time and the description of a reported anomaly were recorded, which permits the measurement of reaction time, vigilance, and information processing for each anomaly. The report time is recorded when a participants clicks on an aircraft to account for the difference in typing proficiency.
- *Global mouse data:* Interaction with hardware is an important measure that can provide insight in the strategies and focus points of participants, as well as activity and workload. Global mouse clicks consists in any clicks recorded on screen, and includes aircraft clicks as well as aircraft label clicks.

- *Commands:* The timestamp and content of all commands issued by participants were recorded, which were used to investigate control strategies and performance.
- *Minimum separation distance:* Measured as a way to quantify performance in terms of safety.
- *Additional track miles:* Measured during the manual control phase of the experiment as a way to quantify efficiency. Represents the difference between track length and the length of a straight line from the entry point of an aircraft in the sector to its designated exit waypoint.
- *Average heading deviation:* A measure to investigate efficiency, it was calculated using Equation 1. The result is a measure that shows the average heading deviation per aircraft in degrees which increases if aircraft are not following the ideal path towards their destination.

$$\Delta \text{ HDGi} = \frac{\text{Sum of heading errors } [deg]}{\text{Number of aircraft in sector}} \tag{1}$$

- *Alert reaction time:* The time elapsed between the moment the alert is shown on screen and the click that dismisses the alert, was measured to investigate vigilance before the failure event.
- *Click rate ratio:* Computed by taking the ratio between the average click rate before automation failure and the average click rate after, as shown in Equation 2. It was used for computing both the global click (all clicks) rate ratio and the label click rate ratio.

$$\text{CRR} = \frac{\text{Number of clicks after failure}/700[s]}{\text{Number of clicks before failure}/4000[s]} \tag{2}$$

- *Questionnaire:* Participants were asked to fill in a questionnaire after the experiment, consisting of several Likert scale and open questions. The Situation Awareness Rating Technique (SART) [26] question set was included, as well as questions about control stategies, order of priorities when supervising automation, experience with and trust in automation and experience with fictional aircraft if present.

Hypotheses

The hypotheses regarding the experiment were mostly expected to be a direct result of the expected increased workload and engagement of operators, leading to higher levels of vigilance and situation awareness.

First of all, it was hypothesised that the implementation of fictional aircraft within an ATC supervisory control environment improves the supervisory control performance of operators, translated through improved anomaly detection rates in the fictional aircraft group compared to the baseline group. This is expected to occur as a consequence of the increased engagement, higher workload and lower boredom level resulting from interacting with fictional aircraft during the supervision task. The first hypothesis is summarised as follows:

HP-1 The implementation of fictional aircraft within a highly automated air traffic control environment improves the anomaly detection rates (i.e., minimise detection misses) of operators.

The positive effect of increased engagement was also hypothesised to translate into enhanced vigilance. Although not measured directly, it is expected that increased vigilance will translate to decreased anomaly report time delays and increased detection rates in participants within the fictional aircraft group, as well as a lower average reaction time for the dismissal of the automation failure alert. Thus, the second hypothesis of the experiment is formulated as follows:

HP-2 The use of fictional aircraft within a highly automated air traffic control environment improves the vigilance levels of operators.

The transition from supervisory to manual control in case of automation failure was hypothesised to benefit from the use of fictional aircraft by operators due to improved situation awareness and vigilance. This should lead to a noticeable improvement in manual control performance among participants in the fictional aircraft group in the immediate period after failure. This hypothesis is formulated as follows:

HP-3 The use of fictional aircraft within a highly automated air traffic control environment improves the immediate manual control performance (safety and efficiency indicators) in case of automation failure.

Finally, it is hypothesised that the mitigation of boredom together with broadening the focus of operators through the means of distributing fictional aircraft throughout the sector will lead to a better traffic overview and increased situation awareness. Thus, the final hypothesis is formulated as follows:

HP-4 The presence of fictional aircraft within an ATC supervisory control environment improves situation awareness of operators.

5. Results

The results obtained from the experiment are presented in the following section. First, the results obtained from objective measures are shown, then the results compiled from subjective measures, and lastly some particular case are discussed in more detail. Due to technical difficulties, detailed mouse position data could only be recorded for Participants 7 to 16. Furthermore, due to the small sample size (eight per group), non-parametric statistical tests were used to analyse the data.

Objective data

The following section presents and compares the results obtained for all participants through objective means during the experiment. Relating to the main hypotheses of the experiment, Mann-Whitney U tests were conducted to determine the effects of the use of fictional aircraft on the short-term (2 minutes after failure) and long-term (12 minutes after failure) heading deviation coefficients, automation failure alert reaction time, and total time delay in reporting anomalies.

Figure 11a portrays the reaction times of participants to the automation failure notification. While the medians for the two groups are the same, the baseline group reaction times show a larger spread. The outlier within the baseline group is a participant who attempted to switch to manual control

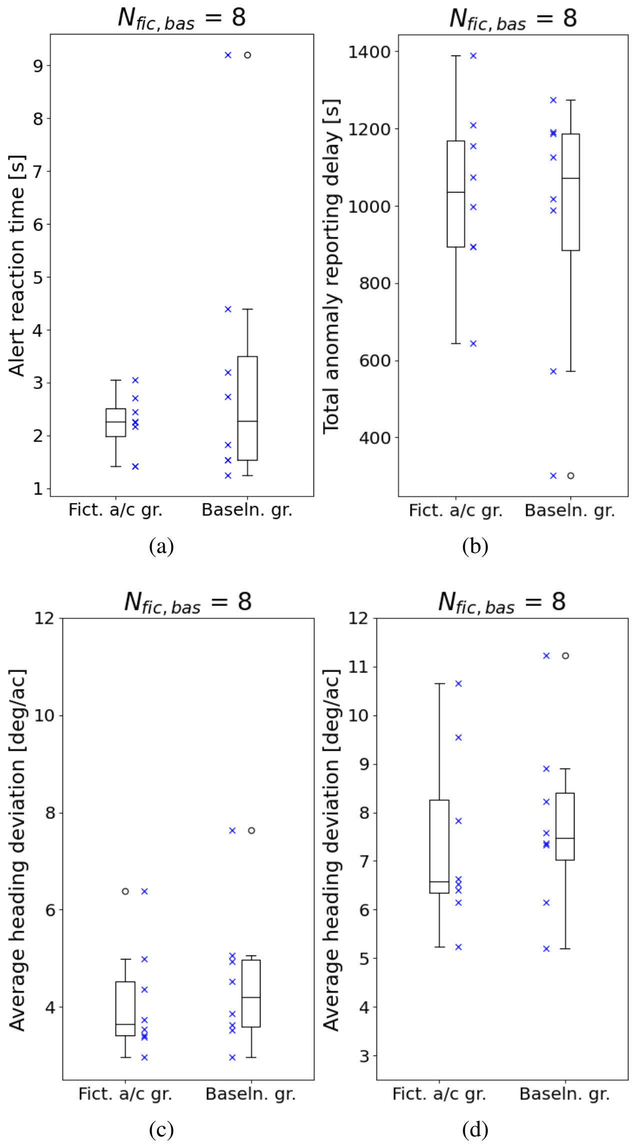


Figure 11: Performance parameters for the two experimental groups: (a) automation failure alert reaction time from appearance on screen to participant dismissal; (b) total anomaly reporting delay with respect to the fastest report time for each anomalous event; (c) average heading deviation over the 700 seconds of manual control; (d) average heading deviation within the first 120 seconds after failure.

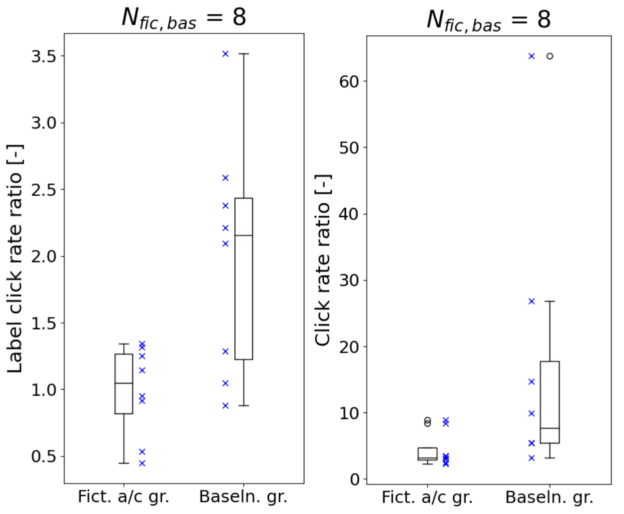
without dismissing the alert, which was not permitted by the software. No significant difference between the fictional aircraft group and the baseline group is observed, confirmed by a Mann-Whitney U test. The same remark can be made then considering the total anomaly reporting delay (the sum of the reporting delay for all anomalies for a participants with respect to the fastest reporting time for each anomaly) presented in Figure 11b. The variance of the dataset is relatively large due to the diversity in participant strategy when reporting anomalies. All participants were told to report anomalies when they were confident it would occur. Thus, the subjective confidence threshold is reflected within the variability of the data set.

The average heading deviation coefficient (Equation 1) does not reveal a significant difference in the data set, although the median of the fictional aircraft group data set is higher both over the whole time interval after automation failure (Figure 11c) and within 2 minutes after failure (Figure 11d). A relatively high variability in the data set can also be observed, as participants employed different strategies when manually controlling aircraft. 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.

Thus, conclusions cannot be drawn regarding hypotheses **HP-2** (fictional aircraft use improves vigilance and reaction times) and **HP-3** (fictional aircraft use improves manual control performance after automation failure). Furthermore, all participants managed to report all the anomalies presented in Table 3, thus no conclusion can be drawn relating to hypothesis **HP-1** (fictional aircraft use improves anomaly detection rates), as no detection misses occurred. An explanation for this comes from the open questions answered by participants after the experiment: most participants from both groups reported that the occurrence of anomalies as well as the beforehand knowledge of their existence contributed positively towards maintaining vigilance, which enabled better anomaly detection performance.

One significant difference between the experimental groups was observed in the ratio between the label click rate before automation failure and the label click rate after the failure event, confirmed through a Mann-Whitney U test ($N = 8$, $U = 12$, $p = .038$). From Figure 12a, it can be seen that the fictional aircraft group has a much lower variability in the label click rate ratio, with a mean closer to 1 when compared to the baseline case. While the absolute number of clicks is a matter of personal strategy, the ratio between the click rates is an indicator of the consistency with which participants interacted with the labels throughout the experiment. This means that participants in the functional aircraft group were more consistent in interacting with labels, whereas participants in the baseline group had on average a much higher click rate after automation failure than before. A more consistent clicking strategy also indicates that participants in the fictional aircraft group on average experienced a more steady transition from manual control to supervisory control.

This trend can also be seen in the global click rate ratio (Mann-Whitney U test, $N = 8$, $U = 10$, $p = .021$), where participants in the fictional aircraft group more consistently interacted with the hardware compared to the baseline group, as seen in Figure 12b. The difference in click rates between the two groups can be seen when the cumulative number of mouse clicks is plotted over time, as presented in Figure 13. The change in mouse activity is visibly more sudden among participants of the baseline group than in case of the fictional aircraft group. The activity of participants seems to be similar in the first 1,000 seconds of the simulation, as fictional aircraft are not initially present on screen. Then the activity of the participants in the fictional aircraft group is higher after this mark as interactions of fictional aircraft and real aircraft become more apparent. Furthermore, there is



(a) (b)

Figure 12: Label click rate (a) and total click rate (b) ratios for the two experimental groups.

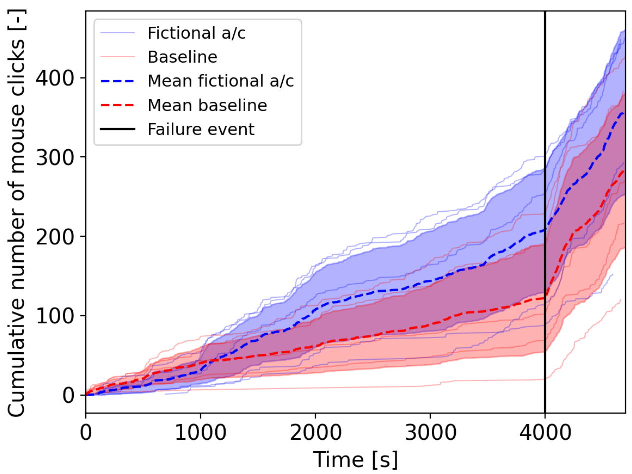


Figure 13: Global mouse clicks as a function of time for Participants 7 to 16, average for each groups with 95% confidence interval.

an initial peak of fictional aircraft between 1,000 and 2,000 seconds, as seen in Figure 9.

More evidence of the benefits of high interactivity with the air traffic comes when comparing the number of label clicks before automation failure with the total anomaly reporting delay, as presented in Figure 14. A monotonic relationship can be observed between the two variables, confirmed by a Kendall Tau-B test (τ_b = -0.393, p = 0.034). Participants that used labels as part of their strategy for supervising real aircraft performed better on average (Participants 1, 10, 12, 15, and 16) when reporting anomalies. This strategy mostly consisted in dragging the labels on top of aircraft to either signify if an aircraft is "clear" or not. The observed trend can be attributed to the increased interaction with the simulation producing increased vigilance in participants. However, definite conclusions cannot be drawn as the anomaly reporting

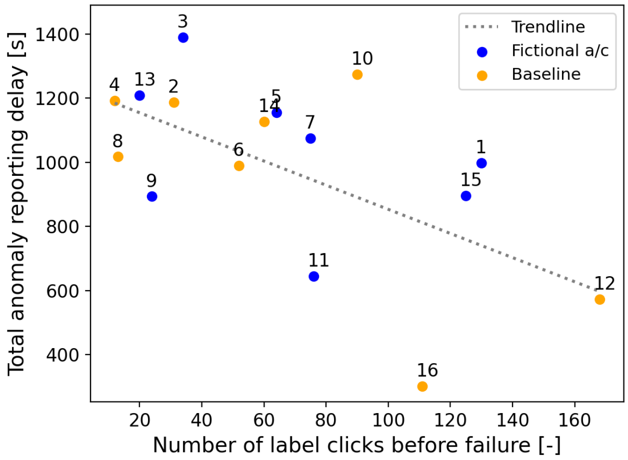


Figure 14: Number of label clicks before failure versus total anomaly reporting delay. Blue points (odd participant numbers) are fictional aircraft group data points, orange (even participant numbers) are baseline group data points.

strategy of each participant influenced the total anomaly reporting delay.

It should be mentioned that all participants detected all anomalies that occurred during the scenario. However, an interesting result can be observed when considering false positive anomaly reports: most reports of this kind were submitted by participants in the fictional aircraft group (five out of eight fictional aircraft participants), whereas only one baseline group participant submitted false positive reports. There were three events that were reported as false positives by participants in both groups:

- **Aircraft CAQ64** experiences a close call with aircraft REL39 and manoeuvres late towards its exit waypoint (reported by Participants 1, 11 and 16)
- **Aircraft NB22J** does not turn towards its exit waypoint when it could, occurs right before the automation failure event, thus most participants have to direct it manually (reported by Participants 1, 11, 15 and 16)
- **Aircraft ZW96D** appears to be close to violating restricted airspace separation requirements (reported by Participants 7 and 13).

While the three mentioned events are not considered anomalies, they do require the attention and supervision by participants during the scenario. The fact that most participants in the fictional aircraft group detected and reported these events could be a result of an overall increase in engagement and situation awareness achieved by the presence of fictional aircraft, resulting in the detection of borderline anomalous events that were mostly not reported by the baseline group. Furthermore, it should be noted that Participant 16, the only person in the baseline group to send a false positive report, reported the highest concentration level among baseline participants. These results support hypothesis **HP-4**, as fictional aircraft group participants achieved higher levels of concentration and situation awareness through the means of interacting with fictional aircraft.

Although not significant towards accepting or rejecting the hypotheses, some particularities in the data reveal important information about the potential effect of automation on controllers, as well as insight into the effects of the experiment design choices. The first instance of particular data can be found when analysing the number of heading commands that participants used after the automation failure event. Figure 15 presents the number of heading commands over time for each participant, as well as the mean average commands for the two experimental groups. The data reveals that two participants in the baseline group had a higher than average command rate than other participants, while no such outliers are present in the fictional aircraft group. Furthermore, it is evident from Figure 15 that the baseline group had on average more commands issued in the initial phase of manual control.

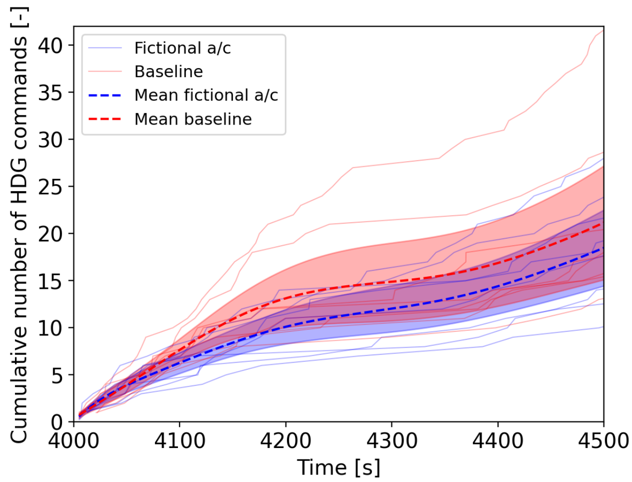


Figure 15: Cumulative number of commands over time for each participant, as well as mean lines with a 95% confidence interval (shaded areas).

This showcases a greater potential problem with the implementation of a high level of automation, and how fictional aircraft could benefit the transition from supervisory control to manual control. If manual aircraft control was simulated before the automation failure event, participants had the opportunity to perform the same task expected after failure, thus mitigating much of the drive for over-commanding aircraft.

Another interesting phenomenon in the anomaly reporting data can be observed when plotting the report time for Anomaly 2, as shown in Figure 16. Regardless of experimental group, participants can be divided in two groups regarding the anomaly report time relative to the fastest reporter: participants that reported the anomaly relatively early, and participants that reported later. Anomaly 2 was the first anomaly that featured an aircraft not turning towards its designated exit waypoint. Therefore, the group division is an indicator for the overall strategy that participants used: some participants chose to report an anomaly as soon as it was noticed, while others elected to wait and observe whether automation would eventually rectify the anomaly.

Figure 16 also shows that reporting strategies were different for other anomalies as well considering the span of the time interval in which participants decided to send a report. For

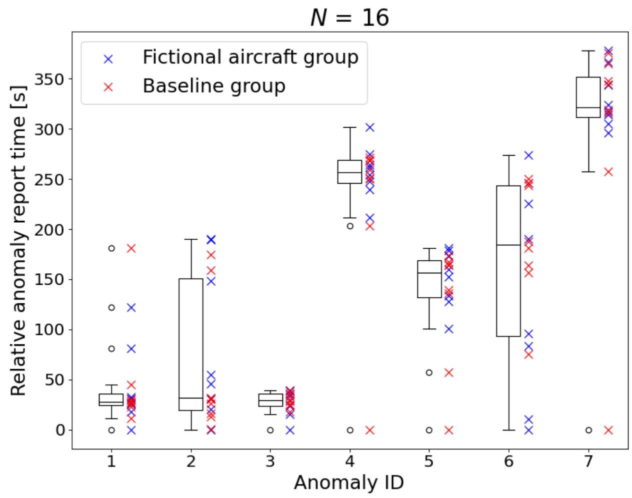


Figure 16: Anomaly report times relative to fastest reporter (data points at report time equal to zero) for all anomalies presented in Table 3.

Anomaly 4 and 7, Participant 16 reported the anomalies correctly more than 200 seconds ahead of other participants, as the participant considered that an aircraft heading towards a restricted area separation breach should be considered an anomaly if the automated algorithm does not correct for it well in advance. This strategy was unique, as other participants preferred to wait in case automation performs later adjustments.

Finally, two participants experienced loss of separation between two aircraft other than the pre-programmed anomalies, both part of the fictional aircraft group. This happened during the manual control phase, approximately 9 minutes after the failure event. One participant described that they were focusing on a different part of the screen, while the other participant reported that they wrongly predicted that intervention is not required. However, as fictional aircraft were not present on screen at the time of occurrence, and a significant amount of time had passed from the failure event, a link cannot be made between the use of fictional aircraft and the loss of separation events.

Subjective data

The following section presents the subjective data obtained from the survey at the end of the experiment session, focusing on the questions on which participants had to answer using a 1-7 Likert scale. Firstly, the SART index computed from the answers to ten questions regarding the supervisory control period of the experiment did not yield a significant difference between the fictional aircraft and the baseline group, which is against **HP-4**. Figure 17 shows that the data is indeed highly variable among the two groups.

However, among the questions of the SART questionnaire, differences can be observed in the following cases: concentration level (*How much did you have to concentrate on the situation? Were you experiencing full concentration (7) or little to no concentration at all (1)?*) and information usefulness (*How much did the received information help with*

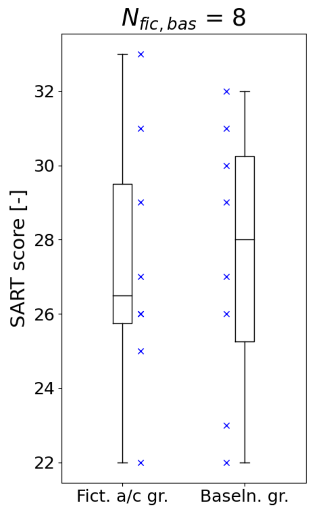


Figure 17: Results of the SART computation for both experimental groups.

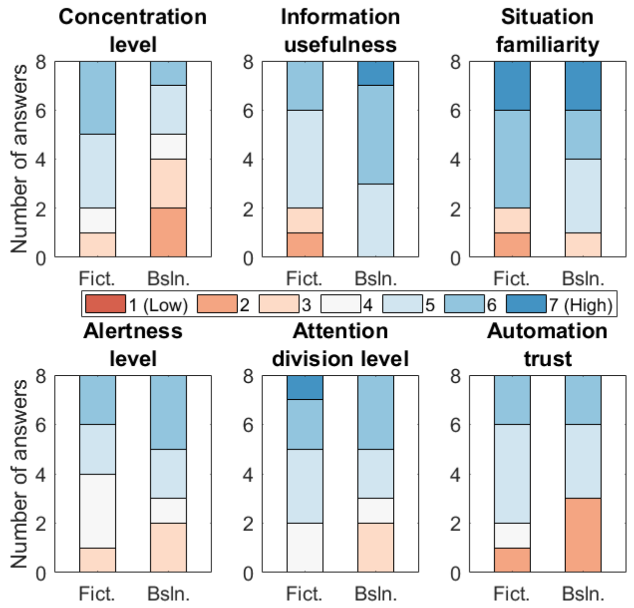


Figure 18: Answers to selected questions represented on a colour-coded Likert scale.

understanding the situation? Have you understood a great deal of knowledge (7) or very little (1)?), presented in Figure 18. In case of concentration level (associated with mental workload), it was expected that, due to the fact that there were more aircraft in total on screen, the fictional aircraft group would experience a higher level of concentration than the baseline group. This difference can be seen in Figure 18, thus supporting hypothesis **HP-4** (fictional aircraft use improves situation awareness), as increasing mental workload was one of the mechanisms through which gamification was expected to be beneficial in mitigating the effects of boredom. A notable data point in the answers to the concentration level question is that one participant (Participant 16) in the baseline group experienced high concentration levels while not having fictional aircraft present on screen. This participant reported a lower automation acceptance level compared to other participants, and therefore experienced increased concentration due to this phenomenon, unique within the baseline group and among the other participants as well, and is considered an outlier.

While not statistically significant, the answers for information usefulness also show a trend: participants in the baseline group indicated that they had a greater understanding of the knowledge received from the display, whereas fictional aircraft group participants scored overall lower. This result shows one of the dangers in overlaying fictional aircraft on top of real aircraft: the screen itself contains more information. However, the information produced by fictional aircraft is not itself considered useful for understanding the situation, thus does not directly contribute to the goal of maintaining the safe and efficient operation of real aircraft. This is backed by some participants in the fictional aircraft group reporting that they found fictional aircraft distracting.

In case of the reported boredom level, one participant in the fictional aircraft group was especially active with controlling fictional aircraft and used them for further training, thus reporting minimal boredom level. However, contrary to the hypotheses, no significant difference was observed in terms of reported boredom levels between the two groups. On the other hand, a trend can be observed in the answers for the attention division level question: participants in the fictional aircraft group overall reported higher levels, which is expected due to there being more elements on the screen that require attention. The answers to other Likert scale questions asked at the end of the experiment did not show significant differences or trends between the two groups.

Finally, each participant was asked whether they found or would have found fictional aircraft beneficial for maintaining vigilance. It is important to recall that all participants, including those in the baseline group, experienced fictional aircraft during training. The answers for this questions are presented in Figure 19. Overall, 5/8 of participants in the fictional aircraft group and 6/8 of participants in the baseline group answered positively to the question. The lower acceptance rate among the fictional aircraft group shows that acceptance rate among the baseline group might be inflated by the lack of prolonged experience with fictional aircraft. However, the difference is relatively small (1/8 participants).

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 in 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 three distinctive strategies of treating fictional aircraft as: (1) real aircraft, (2) as low priority aircraft, and (3) as a game element.

The effect of the treatment of fictional aircraft by participants can be seen in the reaction time for Anomaly 3, due to it being

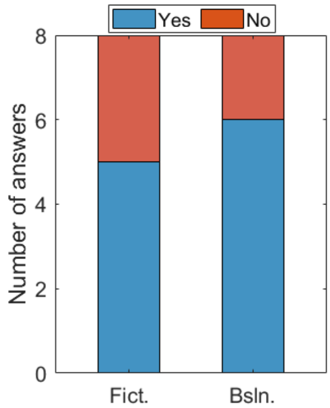


Figure 19: Answers to the question "*Do you think the presence of fictional aircraft helped/would have helped you be vigilant?*".

the only loss of separation event occurring while automation was active. The anomaly can predicted or noticed for a shorter period of time compared to the other anomalies, and is the anomaly with the least variability among participants. This is why, 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 20: lower priority given to fictional aircraft by participants resulted in faster report times for Anomaly 3. This means that participants that 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.

In the open question survey, participants in both groups

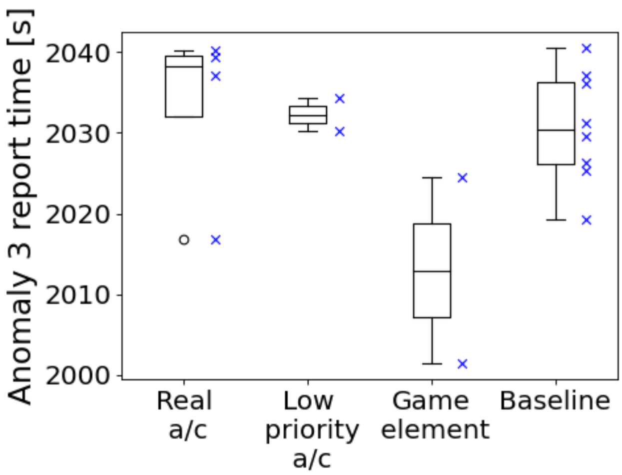


Figure 20: Anomaly 3 reporting time data with fictional aircraft participants divided in three groups depending on their treatment of fictional aircraft: as real aircraft (N = 4), as low priority aircraft (N = 2) and as game elements (N = 2). Baseline group added for comparison.

reported the fact that anomalies helped with maintaining vigilance during the experiment. Thus, even though most participants indicated a high level of boredom during the scenario, 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 research presented in this paper aimed to investigate the effects of the use of gamification within a highly automated ATC environment to enhance controller cognitive abilities when supervising automation. The implementation of gamification was made using fictional aircraft overlaid on the radar screen among automatically controlled real traffic.

The analysis of the objective aggregate results does not reveal a significant difference between the fictional aircraft and the baseline groups in terms of anomaly reporting delay, reaction time and efficiency indicators. Due to the diversity in strategies and technique between participants both in supervisory and manual control, the recorded data has a high degree of variability, thus producing statistically insignificant differences. When compared to luggage screening (field of application of TIP), air traffic control has many more degrees of freedom and changing variables, and thus creates an environment which is more susceptible to differences in personal supervision strategy. However, diversity in behaviour and strategies among participants produced relevant feedback in terms of the perception and acceptance of gamified elements in an ATC context, useful for the further processing and refinement of the fictional aircraft concept. It is expected that professional air traffic controllers would not have provided the same variety in terms of feedback, as they would have had a more homogeneous strategy and perception of fictional aircraft due to their common background and training.

Future iterations of experiments should still consider the use of professional air traffic controllers, as the extensive training and experience means that there will be less ambiguity and variance in how an ATC anomaly is defined and perceived by participants. Furthermore, the concept of fictional aircraft should be extended to be used in the current manual control ATC work environment. The COVID-19 pandemic has shown that situations in which most aircraft are grounded are possible, thus creating a low workload air traffic situation for controllers. Controllers on night shifts could also potentially experience the same low workload situation. In this case, fictional aircraft could be used to increase the concentration and engagement of controllers through the use of a secondary task performed within the work space.

Although most participants reported high levels of boredom, the threshold required for a significant attention and performance decrement was not attained, as all participants reported all intended anomalies 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 as well as less ambiguity and confusion for participants. On the other hand, more extensive training could have achieved the same result, thus making the anomalies more difficult to spot while also lessening confusion.

Overall, experimental results show that the use of fictional aircraft enhances some cognitive processes, and raises required concentration levels. Fictional aircraft also helped participants achieve more consistent interaction with the simulation: most participants in the baseline group were significantly more active in terms of clicks after the failure event compared to before, while participants in the fictional aircraft group showed a very similar level of interactivity. This shows that, from the point of view of activity, participants in the fictional aircraft group experienced a less sudden transition when changing from supervisory to manual control. However, this did not translate into a significant difference in performance between the two groups in the immediate moments after the failure event, which was mostly dependent on the personal manual control strategy of participants.

Another important result of the experiment is given by the false-positive report data. Five out of eight participants in the fictional aircraft 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 is an indication that fictional aircraft helped with maintaining the conflict prediction and situation assessment capabilities of participants.

On the other hand, results also show that the presence of fictional aircraft is not perceived as being useful from the point of view of information flow. Overall, participants in the fictional aircraft group reported that the information they collected from the screen was less useful in understanding the situation. This is an important consideration, as a side effect of gamification is that more information needs to be processed which is not directly useful for the actual goal. In this case, participants were receiving and processing information about fictional aircraft, but the extra information was not useful in supervising real aircraft. Thus, fictional aircraft may mitigate the effects of boredom, but also become a distracting element as mental capacity is directed towards solving fictional conflicts.

One potential issue with the current implementation of fictional aircraft was highlighted by a participant that had undergone air traffic controller training. In the survey, the participant mentioned that fictional aircraft did help to maintain engagement, but in an undesirable way, as issuing one-sided conflict solutions to only one of the aircraft involved is usually inefficient and can induce frustration. Thus, in its current form, the concept of fictional aircraft might be undesirable for use of air traffic controllers due to the one-sided control strategy that needs to be employed. Furthermore, the disappearance of fictional aircraft as the automation failure event occurred leads to a sudden need to re-create the mental image of the traffic present on screen. This sudden change can actually have the opposite effect of what is intended with maintaining the operator in the loop if, within the mental model, the operator does not differentiate between real aircraft and fictional aircraft. This is enforced by the observed trend where participants who treated fictional aircraft as a game element performed better in terms of the report time of Anomaly 3, a loss of separation event which occurs briefly compared to the other anomalous events. Therefore, future developments of the fictional aircraft concept should account for the way these elements are treated, and create a greater separation between what is real and what is fictional to aid controllers in achieving a more robust mental model.

The results obtained in this paper show that the use of fictional aircraft can achieve an improvement in supervisory control performance in monotonous task situations. Gamification can be perceived both positively and negatively, depending on a multitude of factors, including background and personal strategy. Thus, based on the conclusions drawn in this paper, the concept of gamification should be pursued in future research to better understand its potential benefits and discover ways through which the negative effects could be mitigated. Further research is also warranted by the fact that the proposed implementation only tested the basic principles of gamification. Other game elements and strategies should be explored (e.g., scores and achievements) that could be used in a wider range of situations, including normal ATC operations. 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.

7. Conclusion and Recommendations

This research project sought to test the effects of the use of gamification, implemented as fictional aircraft, in a highly automated ATC environment on air traffic controllers. Sixteen students and staff members of 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 aircraft, and had to take over manual control when a predetermined failure event occurred. The baseline group performed the supervisory control tasks without the presence of fictional aircraft on the screen, while the fictional aircraft group had to manually route them through the sector during the task, avoiding conflicts with both types of other aircraft.

The presence of fictional aircraft led to increased reported concentration levels among participants, as well as more consistent interaction with the simulation, which is beneficial for the transition between supervisory and manual control in case of failure. Furthermore, participants in the fictional aircraft group displayed a higher degree of situation awareness as they reported significantly more "close-call" events when compared to the baseline group. However, aggregate objective data for supervisory and manual control performance did not reveal any significant differences between the two groups due to the high variance caused by the personal strategy of each participant.

Future research of the subject of gamification should focus on redesigning the implementation of fictional aircraft such that a more homogeneous treatment of these elements is achieved among operators. Results show that it is beneficial if supervisors are trained to perceive fictional aircraft as a separate game element as opposed to treating them as real aircraft. The experiment also did not achieve a high enough boredom level for producing significant differences in performance due to the occurrence of anomalous events during the supervisory phase of the experiment run. Therefore, future experiments should employ different strategies for measuring supervisory control performance, for example by lengthening the duration of an experiment run and time between anomalies, or by replacing the anomaly reporting time measurement with more advanced techniques such as eye tracking.

REFERENCES

[1] T. Noskievič and J. Kraus. "Air Traffic Control Tools Assessment". In: *MAD - Magazine of Aviation Development* 5.2 (Apr. 15, 2017), pp. 6–10. ISSN: 1805-7578. DOI: 10.14311/MAD.2017.02.01.

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

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

[4] B. Strauch. "Ironies of Automation: Still Unresolved After All These Years". In: *IEEE Transactions on Human-Machine Systems* 48.5 (Oct. 2018), pp. 419–433. ISSN: 2168-2305. DOI: 10.1109/THMS.2017.2732506.

[5] R. Parasuraman, T.B. Sheridan, and C.D. Wickens. "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), pp. 286–297. ISSN: 10834427. DOI: 10.1109/3468.844354.

[6] J. Mercer, A. Gomez, C. Gabets, N. Bienert, T. Edwards, L. Martin, V. Gujral, and J. Homola. "Impact of Automation Support on the Conflict Resolution Task in a Human-in-the-Loop Air Traffic Control Simulation". In: *IFAC-PapersOnLine*. 13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems HMS 2016 49.19 (Jan. 1, 2016), pp. 36–41. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2016.10.458.

[7] V. Pop, E. Stearman, S. Kazi, and F. Durso. "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.

[8] C. Borst, J.M. Flach, and J. Ellerbroek. "Beyond Ecological Interface Design: Lessons From Concerns and Misconceptions". In: *IEEE Transactions on Human-Machine Systems* 45.2 (Apr. 2015), pp. 164–175. ISSN: 2168-2305. DOI: 10.1109/THMS.2014.2364984.

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

[10] S. Nieto and F. Javier. "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. ISSN: 0954-4100. DOI: 10.1177/0954410015596763.

[11] O. Ohneiser, F. De Crescenzio, G. Di Flumeri, J. Kraemer, B. Berberian, S. Bagassi, N. Sciaraffa, P. Aricò, G. Borghini, and F. Babiloni. "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.

[12] C.K. Foroughi, C. Sibley, N.L. Brown, E. Rovira, R. Pak, and J.T. Coyne. "Detecting automation failures in a simulated supervisory control environment". In: *Ergonomics* 62.9 (2019), pp. 1150–1161. ISSN: 0014-0139. DOI: 10.1080/00140139.2019.1629639.

[13] R. Parasuraman, R. Molloy, and I.L. Singh. "Performance Consequences of Automation-Induced 'Complacency'". In: *The International Journal of Aviation Psychology* 3.1 (Jan. 1, 1993), pp. 1–23. ISSN: 1050-8414. DOI: 10.1207/s15327108ijap0301_1.

[14] B. Berberian, O. Ohneiser, F. De Crescenzio, F. Babiloni, G. Di Flumeri, and A. Hasselberg. "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.

[15] E. Paja, M. Poggianella, F. Aydemir, and P. Giorgini. *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, March 19, 2018. 2018.

[16] F. Khatib, F. DiMaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popovi, M. Jaskolski, and D. Baker. "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. ISSN: 1545-9993. DOI: 10.1038/nsmb.2119.

[17] R. Meuter and P. Lacherez. "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. ISSN: 0018-7208. DOI: 10.1177/0018720815616306.

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

[19] F. Hofer and A. Schwaninger. "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.

[20] T. Kim and K. Werbach. "More than just a game: ethical issues in gamification". In: *Ethics and Information Technology* 18.2 (June 2016), pp. 157–173. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/s10676-016-9401-5.

[21] C. Borst, R.M. Visser, M.M. van Paassen, and M. Mulder. "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. ISSN: 2168-2305. DOI: 10.1109/THMS.2019.2919742.

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

[23] *MUAC implements additional sectors in Netherlands airspace to meet increasing traffic demand*. URL: https://www.eurocontrol.int/news/muac-implements-additional-sectors-netherlands-airspace-meet-increasing-traffic-demand (visited on 2020-10-14).

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

[25] P.A. Hancock. "In Search of Vigilance The Problem of Iatrogenically Created Psychological Phenomena". In: *The American psychologist* 68 (Oct. 22, 2012). DOI: 10.1037/a0030214.

[26] EUROCONTROL. *Situation Awareness Rating Technique (SART) — HP repository*. 2012. URL: https://ext.eurocontrol.int/ehp/?q=node/1608 (visited on 2020-04-24).

II

Preliminary Report

Previously graded under AE4020

1

Introduction

The report at hand presents the literature study and preliminary analysis conducted within a research project focused on issues arising from the implementation of a high degree of automation in the air traffic management industry. The scope of the project is to produce a tool that would aid air traffic controllers in improving and maintaining their supervisory control abilities in monotonous work environments.

The report is structured as follows: the first part focuses on exploring the background and present context of the demand for increased automation in the field of air traffic control. This is done through the display of the research goals and methods, as well as the presentation of the results of a literature study. The second part of the report is centred around the research and design choices made for the development of both a cognitive aid tool and an experimental setup to test it.

1.1. Research Context

Air traffic is predicted to continue growing in the upcoming years, with estimates for the increase in yearly passengers between 70% and 185% until the year 2070 [1]. Even with the SARS-CoV-2 pandemic of 2020, which heavily affected the aviation industry and thus these estimates, future demand for increased capacity, efficiency, and safety should be accounted for. In order to accommodate this, the framework and infrastructure of the support facilities of the air traffic control industry need to continue evolving in order to achieve more efficient, safe and productive operations within densely populated air sectors. To attain this, several proposed solutions are currently being researched, revolving around different degrees of automation and free route implementation.

The 2020 European Air Traffic Management (ATM) Master Plan [2] states that a high level of automation is sought to be implemented with the scope of accommodating the predicted traffic growth, as well as improving safety and efficiency of air transportation. This will be facilitated by the integration of information systems through the implementation of a digital data-link between ground control and aircraft, which will enable ground-based computers to transmit information directly to the pilot and the on-board computer. Furthermore, there is also a desire to implement free route concepts, shifting some of the responsibilities that are currently being handled by ground-based air traffic control (ATC) to the cockpit.

However, this entails that most tasks of a current air traffic controller (ATCo) will be automated or relocated. While the degree of automation implementation in ATM is a highly debated subject, operators will potentially assume the role of system supervisors for the majority of tasks. This has great implications within the work environment of an ATCo, as the control task fundamentally changes. Most research shows that a high degree of automation has a negative effect on human cognitive abilities. Metzger and Parasuraman [3] found that the performance of controllers is affected when executing a “management by exception” task, influenced by factors such as task monotony or loss of vigilance and situation awareness. Volz et al. [4] found that skill degradation is a consequence of high degrees of automation, as less time is spent practising manual control skills.

Reliable automation also introduces complacency and overconfidence in the supervisors [5]. As automation does have the potential to fail given enough time, these effects can affect the detection of such events and the performance of operators after manual takeover. The industry is in agreement that the unaided management by exception (i.e., the system supervisor only intervenes if a mishap occurs in automation) of air traffic is not a viable option [2], therefore research is oriented towards the development of tools and strategies for improving and supporting operators performing supervisory tasks.

Some solutions have been developed and implemented within the aviation industry, for example in the luggage screening domain. In this work environment, operators must be vigilant for extensive amounts of time, while exceptions (such as prohibited items) occur very rarely. To mitigate the negative effects of this, several airports have implemented the concepts of threat image projection (TIP): the use of simulated threats within luggage to maintain operator vigilance and situation awareness [6].

The research presented in this report explores the use of threat image projection within air traffic management as a vigilance and situation awareness enhancement strategy. The proposed ATC system augmentation consists of the simulation of additional aircraft in the airspace an operator supervises that need to be manually controlled. The simulated aircraft would be overlaid on top of real traffic, which is fully controlled by automated systems, and thus are unaffected by the simulated traffic. The task of the operator is to guide the simulated aircraft through the sector and resolve conflicts that might arise with other aircraft, both real and fictional. This would mean that the benefits of using automation are achieved (increased safety and efficiency) while also mitigating the negative effects (boredom, lack of engagement). It should be noted that the research project at hand seeks to explore the use of gamification at a conceptual level, and does not seek to present a final ATC tool design.

1.2. Research Approach

The following section presents the methodology through which the proposed system augmentation concept will be researched, as well as the hypotheses, the guiding research questions and research aim.

As previously described, the concept that is explored consists of overlaying simulated aircraft on top of real traffic, the latter being controlled by automated systems. This is an example of gamification, the concept of implementing game elements within non-gaming contexts, which has gained traction as a method to enhance cognitive abilities in humans. To summarise, the aim of the research project at hand is to achieve improved manual and supervisory control performance of air traffic controllers by means of implementing gamification concepts in a fully automated ATC environment.

The main research question is thus focused on the effect of using gamification as an engagement, vigilance and situational awareness enhancement tool as a way to improve supervisory control performance:

"**How does the use of gamification concepts, in the form of simulated aircraft conflicts, affect operator supervisory and manual control performance of real traffic after automation failure in a highly automated upper area ATC control environment?**"

Several sub-questions can be formulated for each part of the research project based on the main research question. The following questions are sought to be answered through the execution of a literature survey:

SQ-1 What is the effect of the implementation of a high degree of automation in an ATC environment on controllers?

- **SQ-1.1** What elements of the ATC environment are predicted to be automated in the next thirty years?
- **SQ-1.2** What human factors are influenced by the use of a high degree of automation?
- **SQ-1.3** What are the design considerations that influence human cognitive factors in a highly automated ATC work environment?

The following questions are answered within the preliminary analysis and design part of this research project:

SQ-2 What are the options found within the field of gamification to mitigate of the negative effects of a highly automated work environment on human cognitive abilities?

- **SQ-2.1** What are the effects of the gamification of workplaces on human cognitive abilities?

SQ-2.2 What are the ethical considerations of the implementation of gamification concepts in work environments?

SQ-2.3 What gamification concepts are viable for implementation within a highly automated ATC work environment?

SQ-2.4 What fictional aircraft traffic pattern has the potential to yield the best results in terms of operator vigilance and situational awareness?

SQ-2.5 What visual and auditory display elements have the potential to yield the best results in terms of operator vigilance and situational awareness?

And finally, the following set of sub-questions will be answered through the development and execution of an experiment:

SQ-3 How does the proposed aid tool affect the supervisory and manual control performance of controllers in a highly automated ATC environment?

SQ-3.1 How does the proposed supervisory control aid tool affect situational awareness in operators when compared to an unaided control scenario?

SQ-3.2 What is the influence of the number of fictional aircraft on operator supervisory and manual control efficiency and safety when an intervention is required after automation failure?

SQ-3.3 What is the effect of the proposed tool on the reaction time of system operators in case of automation failure compared to an unaided supervisory control scenario?

The research questions are related to dependent measures that can be measured in an experimental setup: workload, reaction time, situational awareness and other control performance parameters (e.g., track miles, number of conflicts, etc.).

The main hypothesis is that, as the operator is given an active task within the work space, situation awareness and vigilance are improved by maintaining engagement. This would in turn improve failure detection time and control performance after a failure event. More predictions on the potential results of the project are made and discussed in Chapter 5.

To achieve the research goal, a literature study is performed according to the research map presented in Figure 1.1. This includes exploring the current capabilities of ATC systems as well as the projected requirements for future ATC systems, which can highlight the requirements that the proposed system must fulfil. The effects of the implementation of automation on humans and on the ATM environment are also explored to gain full understanding of the issue that will potentially affect the future of aviation. In order to implement the concepts of threat image projection and gamification within ATC, their current implementations in their respective domains are investigated. This could help with predicting their impact on controllers and predict the results of using them. As an experiment will be performed to investigate the hypotheses, research is also done into their development and execution, especially in the domain of air traffic control.

As previously mentioned, an experiment is required in order to test the hypotheses proposed in this research project. The second phase is therefore the development of a simulation environment in which the concept of threat image projection can be implemented. Together with performing experimental design, this phase will end with the execution of the experiment. The third and final phase consists in the analysis of data, which will lead to the accepting or rejection of the hypotheses.

1.3. Report Structure

The report is divided into two parts: Part I presents the research background and context in Chapter 1, and continues with the literature study performed for acquiring knowledge about the research topic in Chapter 2. Part II starts with the presentation of the preliminary design of the proposed tool in Chapter 3, includes the experimental hypotheses and setup, described in Chapter 4. Chapter 5 presents the conclusions of the entire report.

Figure 1.1: Research map used for the literature study.

2

Literature Survey

The following chapter presents a summary of the literature survey conducted with the scope of acquiring information about the potential future developments in ATC automation, including the effects on controllers and the work environment. As concepts that could potentially solve these problems, threat image projection and gamification were researched to obtain a better understanding on the benefits of their implementation, as well as the shortcomings. Finally, display design considerations are explored as to produce an ecological interface for operators.

2.1. Automation in Air Traffic Control

Air traffic is predicted to increase significantly in the near future [7], with the current ATM framework not being able to accommodate such growth. A report produced by SESAR [8] shows that, by the year 2030, the delay and congestion within current air sectors will overtake the levels reached during the Kosovo crisis in 1999. A widely accepted solution within the industry is the adoption of automation in a wider range of processes and activities within ATM. According to the European ATM Master Plan [2], automation is expected to play a heavy role in the air traffic control process, and achieving a high level of automation is one of the goals of the SESAR project.

2.1.1. Levels of Automation Taxonomy

The term “automation” can refer to a broad range of applications within ATM, from data management to conflict detection and resolution. In order to describe the degree of automation used in a system, several taxonomies have been defined over time. A popular classification is given by Sheridan [9], however it is mostly focused on the decision making process. Within

SESAR, the level of automation taxonomy used is based on the standard used in on-road vehicles [10], presented in Table 2.1.

Table 2.1: Levels of automation taxonomy used by SESAR [2], with the degree of automation for each task and level being rated on a scale from 1 (little to no automation) to 4 (complete automation).

Level	Definition	Information acquisition and exchange	Information analysis	Decision and action selection	Action implementation	Autonomy
0	Low automation	IV	III	I	I	I
1	Decision support	IV	IV	II	I	I
2	Task execution support	IV	IV	IV	II	I
3	Conditional automation	IV	IV	IV	III	II
4	High automation	IV	IV	IV	IV	III
5	Full automation	IV	IV	IV	IV	IV

The levels of automation in the context of air traffic control can be defined in the following way:

Level 0 Automation mainly supports controllers through aircraft information processing and analysis;

Level 1 Automation supports controllers in the conflict resolution decision making process;

Level 2 Automation provides conflict resolution advisories to controllers, which approve or dismiss them;

Level 3 Automation initiates simple tasks; can be adaptive to the needs of the controller;

Level 4 Automation initiates most tasks, lacks complete autonomy (controller supervises);

Level 5 Automation performs all tasks, has complete autonomy.

2.1.2. Current Use of Automation in Air Traffic Control

Automation is currently employed in the air traffic control process in several functions, such as collision detection and alerting, trajectory prediction and information processing [11]. While tools are used by operators as decision support, the decisions are still made by humans and not by machines. According to the taxonomy presented in Table 2.1, this is equivalent to Level 1 automation. According to Noskievič [12], the main tools presently being used by air traffic controllers are:

- **Trajectory prediction:** the use of historical data for path estimation and optimisation;
- **Medium-term conflict detection:** (MTCD) detects conflicts up to 15 minutes in advance;
- **Short-term conflict alert:** (STCA) detects conflicts up to 2 minutes in advance;

- **Arrival and departure traffic manager:** used to optimise runway capacity;
- **Minimum safe altitude warning:** uses trajectory prediction and terrain data to warn controllers of inadequate altitude;
- **Area proximity warning:** alert when an unauthorised aircraft enters restricted airspace.

The implementation of such information acquisition and analysis tools has improved the performance of air traffic management centres, such as an 11% increase in the German upper airspace capacity after the implementation of a medium-term conflict detection tool [13]. This shows that there is potential for obtaining better air traffic control performance if automation is used, therefore SESAR has set the goal to reach Level 4 ATC automation by the end of the year 2040 [2].

2.1.3. Future Implementation of Automation in ATC

The benefits of a high level of automation on the efficiency and safety of operations have been extensively researched. Nieto and Javier [14] describe that the aviation industry will reach level "4" or "5" of automation by the year 2050, which will increase the airspace capacity. This vision is also employed by the SESAR project [2].

The use of automation will also further permit the implementation of free-flight concepts within the European airspace, which have been proven to be beneficial for the capacity of an air sector while improving safety [15]. This further implies that, regardless of which agent will perform conflict resolution, the trend is to shift the decision making process away from ground human operators. This conclusion has also been expressed within SESAR through the STRESS [16] and MINIMA [17] projects.

A high level of automation entails that the air traffic controllers assume the role of a system supervisor, managing the automated processes through exception [18]. Even through being proven as beneficial for the efficiency and safety of controlling air traffic, resistance might be encountered within the industry. Issues arise from two points of view: the tendency of air traffic controllers to reject automation [19], and the negative impact of automation on the cognitive performance of human operators [20].

2.1.4. Effect of Automation on Air Traffic Controllers

With increasing levels of automation, the performance and work environment of an air traffic controller is affected. Extensive research has been conducted on what the effects of implementing automation are on cognitive performance parameters. The ironies of automation [21] predict several problems with the replacement of humans, including skill degradation, vigilance deterioration, and reduced human engagement. In later studies, these issues have been experimentally proven.

Thackray [22] experimentally proved that human operators are not suitable for supervisory control tasks if unaided. More recent experiments, such as the one presented in Foroughi et al [23], performed with an updated controller setup, obtained the same findings in an environment where automation reliability is relatively low. This enforces a conclusion presented in the European ATM Master Plan [2, p. 87], stated in the context of task allocation between air traffic controllers and the system (machinery, automation) in the future:

"The traditional belief that the human will manage unexpected events unaided or unsupported is no longer viable."

This shows that the air traffic management industry is moving towards the implementation of automation within the work space of air traffic controllers. However, this has the potential to negatively influence cognitive factors in human operators if not implemented carefully.

Parasuraman et al. [24] describe a wide range of issues that could arise from automating ATC tasks. The article mentions several factors that are influenced by automation as well as recommendations on the level of automation that should be employed in ATC, presented in Figure 2.1. The level of decision automation equivalent with the recommendation is between Level 2 and 3, which is not compatible with the goals set by SESAR (Level 4-5). This could imply that knowledge of the capabilities of automation has improved, with the SESAR project implying that the negative effects of its implementation will be mitigated.

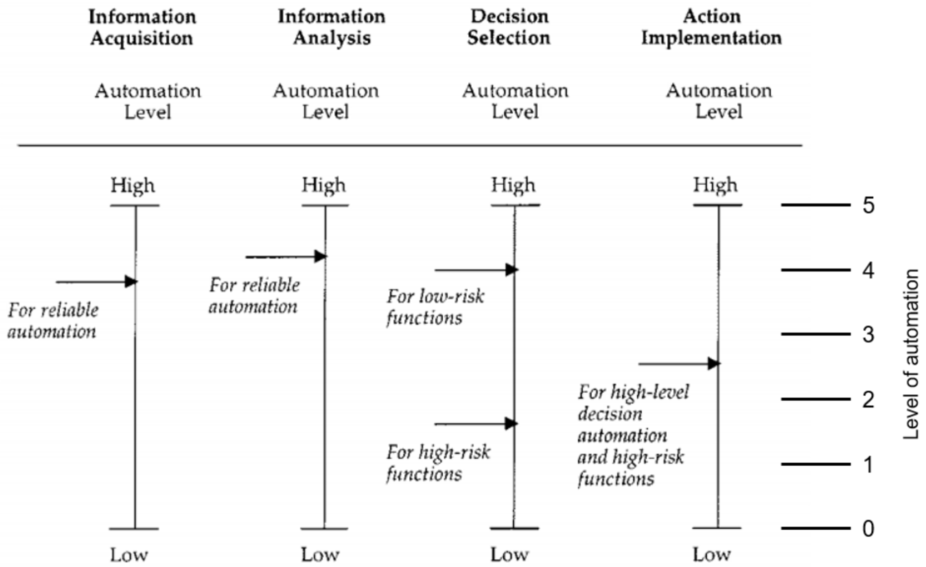


Figure 2.1: The automation level recommended by Parasuraman et al. [24] for ATC tasks.

It is therefore necessary to identify and account for the effects that changes in the work environment produce in humans. From literature, several human cognitive factors are identified which are influenced by the implementation of automation. These will be presented and analysed below.

Situation Awareness

Situation awareness (SA) is defined, according to Endsley [25], as "the perception of elements in the environment, the comprehension of their meaning and the projection of their status in the near future". Endsley [26] divides the extent of situation awareness into three levels:

Level 1 Perception of elements in the environment;

Level 2 Understanding the current situation;

Level 3 Extrapolation of the future state of the system.

A Level 1 situation awareness means that the operator perceives the changes that occur in the environment, but does not interpret them. The next level means that the operator is both aware of the environment and understands the events that occur within the workspace. However, the ability to extrapolate using the current events only comes with the highest level of situation awareness. The ATCo can be aided to attain these levels through the use of automation (information processing automation, aircraft separation monitors and prediction aides).

Situation awareness is a critical cognitive state for a supervisory controller, as it is heavily influenced by the implementation of automation. Due to the effects of a lack of engagement and monotony of a highly automated work environment, humans tend to experience a decrease in attention and focus, thus lowering situation awareness [27].

Parasuraman et al. [28] show that situation awareness, mental workload and trust are constructs that can be used to predict human-system performance. This implies that enhancing SA does lead to an increase in performance. There are several ways to achieve this, mostly related to maintaining focus and operator engagement.

Mercer et al. [29] demonstrated through an experiment which sought to study the effect of conflict detection automation in air traffic control that operators that performed the detection task unaided tools were more engaged, had better situation awareness, and therefore resolved conflicts in shorter times if they were involved in the conflict detection task. Nunes [30] experimentally determined that display augmentations can be used to enhance the situation awareness of users in a free-flight air traffic control scenario.

Therefore, situation awareness is established in literature as an indicator for human performance in a supervisory control environment, and as a factor that can be improved through work domain augmentations. Two types of failures can occur from low situation awareness, documented by the MINIMA project [17], presented below.

- **Failure to detect:** occurs on the lowest level of situation awareness (level 1). Leads to between aircraft conflict detection errors or delays in air traffic controllers, and can occur if controllers are passively monitoring automation [31].
- **Failure to understand:** occurs on higher levels of situation awareness (levels 2 and 3). According to Berberian et al. [17], operators experience errors in understanding and thus

overcoming the problematic situation. Furthermore, operators experience “automation surprises”, showing that they show a lack of understanding over the decisions taken by automation.

Mental Workload

Before concepts such as situation awareness were explored, mental workload was for a long time the main subject of interest when considering human-machine interaction. The effect of automation on mental workload can be a significant one, either reducing or increasing it depending on the implementation [3]. The implementation of current assist tools within air traffic control aim to lower controller workload by assisting with tasks such as data acquisition and processing.

However, a high level of automation can also negatively impact human mental workload [32], and therefore decrease supervisory control performance. Therefore, the assumption that automation leads to lesser mental workload (regardless of the level of automation) does not always stand, as concluded by Parasuraman et al. [20], thus tools need to be developed such that mental workload is kept within boundaries that ensure best controller performance.

Vigilance

Vigilance is defined in Warm et al. [33] as the ability to maintain focus for extended periods of time. It is therefore an important factor in the context of supervisory control, where this ability can be affected by the lack of involvement in the control task of the operator [20], producing a state of boredom.

Vigilance has been long established to be correlated with performance parameters such as reaction time, as demonstrated by Buck [34]. Berberian et al. [17] mention that a change in vigilance level causes negative effects on the performance in failure detection and situation awareness of the system.

Therefore, as with situation awareness, automation should be implemented in such a way as to maintain the ability to be vigilant in the humans which supervise it. One method to do this is to maintain engagement, as shown experimentally by Pop et al. [35].

Complacency

A major negative effect of the implementation of a high degree of automation is that operators might become over-reliant and complacent, especially if the reliability is high. This was determined by Parasuraman et al. [36] through a series of experiments, thus affecting the ability of humans to detect automation failures.

Wickens et al. [37] show that the way the aid fails has a major impact on the manifestation of complacency: automation making wrong decisions is worse than automation failing completely (and therefore completely deactivated) from the point of view of intervention performance. Furthermore, ensuring that the operators are aware of the shortcomings of automation lessens the effect of complacency.

Trust

Trust in automation has gained traction as a factor to be considered when studying human machine interaction. As mentioned previously, Parasuraman et al. [28] argues that trust can be used as an indicator for human-machine interaction. However, trust is not as straightforward as the other presented cognitive factors: both high and low trust in automation can be detrimental to the performance of an operator fulfilling a supervisory role.

Lee and Moray [38] determined that trust increases as controllers become more familiar with the system, and is correlated with better cooperation, and thus increased performance of a human-machine system. However, high levels of trust lead to complacency and low failure detection performance, as mentioned earlier. On the other hand, Dixon and Wickens [5] show that low reliability automation, which induces distrust, yields higher workload for operators than high reliability automation. Therefore, trust in automation must be kept at a level that benefits human-machine systems the most: high, but not excessive.

Cognitive Tunnelling

Cognitive tunnelling is defined by Wickens [39] as a phenomenon that occurs when an operator focuses on certain parts of the work space for an extended period of time, resulting in the neglect of other events. This can occur regardless of the presence of automated tools, however these tools can potentially cause this phenomenon if the controller focuses excessively on them.

Shorrock [40] mentions that cognitive tunnelling occurs more in highly stressful and workload intensive situations. In case of air traffic control, a dense air space could lead to failure to detect conflicts in a timely and safe manner. Short and medium-term conflict detection tools have been implemented to prevent this by increasing the situation awareness of the human operator. However, future automation needs to be designed such that a high level of situation awareness is maintained, thus reducing the effect of cognitive tunnelling.

Skill Degradation

With the increasing level of automation in all domains, human operators are becoming system supervisors. One of the downsides of this is skill degradation, which occurs when humans do not perform hands-on control for an extended period of time, and therefore experience deterioration in their ability to perform manual control tasks. This phenomenon has been noticed in pilots as the cockpit became increasingly automated. Haslbeck and Hoermann [41] conducted a study, showing that skill degradation occurs especially in long-haul flights, when pilots perform less manual control.

In the case of air traffic control, skill degradation could become a major problem with increasing automation, as controllers rely on skill-based behaviour to detect and resolve conflicts. This could pose a problem if manual intervention is required when automation experiences failure, as it could lead to decreased control performance. Kirwan [42] determined that air traffic controllers could be affected by the lack of practice in the following tasks: state prediction, evaluation, decision making and optimisation of the airspace. Therefore, future tools need to be developed to account for this phenomenon and ensure its mitigation.

2.1.5. Automation Acceptance and Reliability

Automation has been difficult to implement in air traffic management, as controllers are selective with the tools they are willing to use [43]. However, as mentioned previously, most literature agrees that automating controller tasks is the most probable solution to future capacity constraints. By conducting a survey among air traffic controllers, Bekier [19] determined that the line between automation acceptance and rejection is drawn at the decision making process: controllers want to be in charge of making decisions. This vision is incompatible with the vision presented in the European ATM Master Plan 2020 [2], which aims to achieve a high level of automation (Level 4 by the year 2040) within ATC.

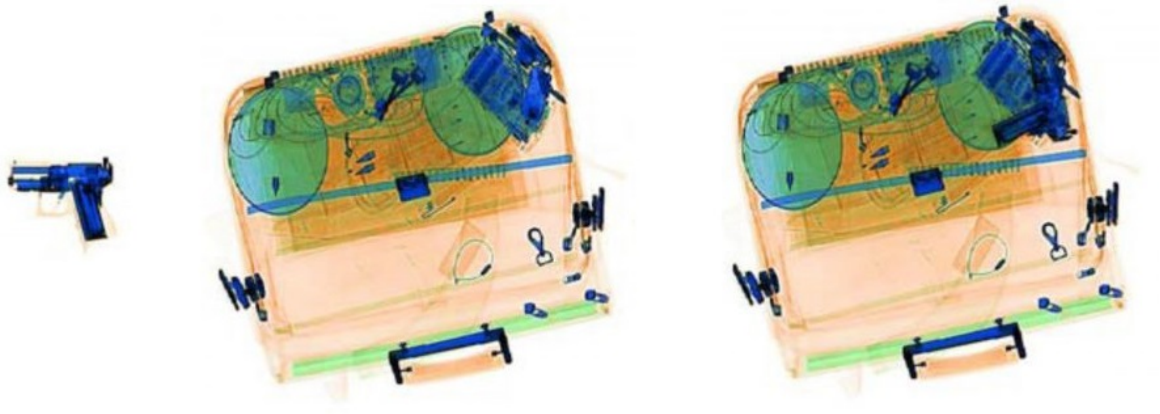
Automation acceptance could improve if factors such as trust and job satisfaction are improved [19]. In order to improve trust, the high reliability of automation needs to be proven. Metzger and Parasuraman [3] determined that, while highly reliable automation leads to an increase in conflict detection and resolution performance, unreliable automation worsens performance with respect to manual control, as the trust in automation is low. Rovira and Parasuraman [27] determined that even relatively reliable automated conflict detection negatively affects controller performance if it is imperfect. As no system is perfectly reliable, trust will always play a factor in the implementation and use of automation. Therefore, future air traffic control automation will have to be highly reliable in order to achieve better human-machine performance.

Another method through which automation acceptance could be improved among air traffic controllers is adaptive automation. Ohneiser et al. [18] states that automation could only be used if cognitive parameters such as workload and focus drop below a certain level, with varying automation intensity. This would mean that automation is mostly a safety net in case the human operator cannot cope with high density airspace scenarios. Air traffic controllers would therefore retain their decision making role, with automation intervening only when necessary.

Lately, the field of artificial intelligence has been gaining traction and infiltrating many other domains. Westin et al. [44] proposes that automation acceptance by air traffic controllers could be improved if the automation algorithms are tailored to the control strategy of the supervisor. This would mean that automation would have to present expected behaviour, and thus be more predictable, leading to lower workload for its supervision.

2.2. Threat Image Projection

As previously discussed, the loss of situational awareness is a factor that can affect both manual and supervisory performance for a human operator. Many concepts have been proposed and implemented to combat this effect, ranging from the operational domain (e.g., maintaining



(a) (b) (c)

Figure 2.2: Example of threat image projection implementation [46]: (a) is the fictional threat, (b) is the screened baggage, (c) is the image the operator must examine. The threat is overlaid in the top right corner of image (c).

pilot engagement by turning some tasks over to be performed manually) to more technically oriented solutions (e.g., adaptive automation). However, a potential solution that could find applicability in the air traffic management industry has been implemented in other domains.

Baggage screening is considered a highly monotonous activity performed widely across the world. Operators must watch a screen display the x-ray scans of passenger bags, and notice if there are any prohibited items (exceptions). This activity can be compared with an operator performing management by exception supervisory control, as passengers are expected to comply with the rules and not bring prohibited items on aircraft. However, the similarity extends to the cognitive challenges that humans experience in a supervisory control environment: loss of vigilance and boredom. Meuter and Lacherez [45] experimentally determined that longer screening operator shifts lead to a higher probability for threat detection errors (level 1 SA failure).

To mitigate this, many airports have implemented the concept of threat image projection (TIP) within the workplace of baggage screening. As described by Schwaninger [6], TIP consists in projecting fictional images of threats on the screen, and overlay them on top of regular baggage contents, as presented in Figure 2.2, where a firearm shaped fictional image (a) is overlaid in the top right corner of the scanned baggage (b) to produce the final image (c). The threats originate from a large database, containing a wide range of objects. If the operator detects the exception, they are informed that the threat was fictional, which is then removed from the screen. The decision time is also limited: if the operator fails to detect the conflict in a given amount of time, then they receive a “miss” notification.

By increasing the overall threat frequency, an increase in detection performance has been observed by Hofer and Schwaninger [47] when threat image projection has been used in combination with a large database of threats. On the other hand, a limited threat database resulted in lower detection performance, as operators became more familiar with the objects. Therefore, variety is an important factor in the deployment of threat image projection, and is achieved using three parameters, explained below.

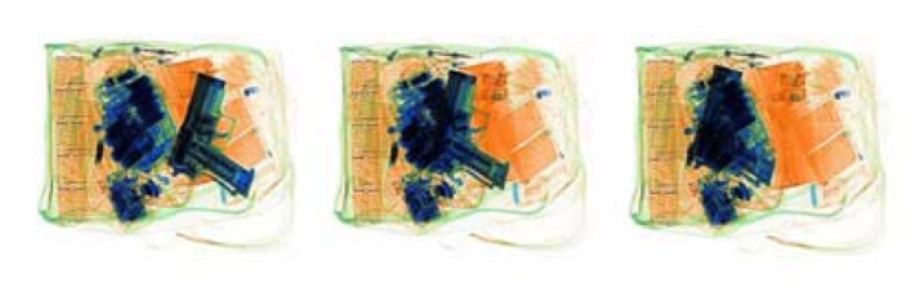
- **View difficulty:** The threat object can be presented to the operator in differently configured views, as portrayed in Figure 2.3. The third view (Figure 2.3c) of the firearm offers little information with respect to the shape of the object, and thus might be misinterpreted and missed by more inexperienced operators. This parameter can be compared with the difficulty of detection and resolving a conflict in air traffic control originating from the positioning of the two conflicting aircraft (i.e., conflict angle, aircraft velocities, aircraft positioning).
- **Superposition levels:** The superposition location of the threat on the scanned baggage can alter the difficulty level, as seen in Figure 2.4. If the contrast between the threat and the other objects in the image is high, then the operator can more easily detect it. An equivalent parameter within air traffic control would be the number of aircraft with which a single aircraft is in conflict with at a given time: the operator would focus on the more obvious conflicts, and fail to detect or account for the less imminent ones.
- **View complexity:** The density of objects in a bag influences the experienced workload for the screening operator: if more objects need to be analysed, workload increases, as portrayed in Figure 2.5. Thus, varying the view complexity will affect the threat detection difficulty. The view complexity parameter can be compared with the air sector aircraft density: more aircraft in a sector can lead to a lower probability of conflict detection as the attention of the operator is distributed.

Threat image projection also has the potential to be used for screening operator performance assessment, as the amount of object detected or missed provide information about the alertness and vigilance of screening operators. With the increased frequency of threats to be detected, and the instant feedback that is given to the operator, managers can use the tool to improve performance and the work environment.

Threat image projection is therefore a relatively new concept that could potentially be implemented with the air traffic control domain. According to Schwaninger [6], the use of TIP improves the motivation and vigilance of screening operators. Thus, as the transition to a high degree of automation in ATC is inevitable, the implementation of this concept should be explored, especially as it has never before been attempted.

(a) (b) (c)

Figure 2.3: Implementation of threat image projection in baggage scanning: view difficulty [48]. From left to right, the object is more difficult to identify as a firearm due to the point of view.



(a) (b) (c)

Figure 2.4: Implementation of threat image projection in baggage scanning: superposition levels[48]. From left to right, the firearm is increasingly obstructed by other objects in the luggage.

(a) (b) (c)

Figure 2.5: Implementation of threat image projection in baggage scanning: view complexity[48]. From left to right, there are more objects to account for in the bag.

2.3. Gamification

Gamification is a relatively new concept that has gained traction as a management technique. While there is no established definition, it implies the use of game principles outside of game contexts [49]. While it is used in many domains spanning a large range of uses, it has the potential to be used as a motivation, vigilance and situation awareness enhancement technique in the context of air traffic management.

2.3.1. Implementations of Gamification

Gamification is currently employed in a wide range of domains, from management techniques (for example, the PACAS project [50], which developed a gamified change management process for ATM) to applications in virus research: scientists used mobile games to have users create simulations of the behaviour of the HIV virus [51]. These implementations seek different goals with the use of gamification: the first case focuses on collaboration, the latter on motivation to contribute to research. This shows the flexibility of the use of gamification concepts.

Threat image projection could also be considered an example of a current implementation of gamification in airport x-ray baggage screening, as operators are given active feedback on their performance, but are also scored according to their performance. While threat image projection is a promising concept that could potentially achieve increased engagement in operators, it could potentially contribute to the degradation of motivation, as the “stakes” become lower with the implementation of fictional threats. This is especially the case if the operators are aware that the threat they are facing is fictional, as is the case with the concept explored in this research project. However, this can potentially be mitigated by motivating operators through other methods, such as rewards or methods that enhance internal motivation.

Gamification can be implemented through the use of so called “game mechanics”, such as [49]:

- **Points:** common gamification technique, can be used for direct feedback or for its motivational value;
- **Levels:** give players a sense of progression, which can maintain engagement and motivation;
- **Leaderboards:** can be used to stimulate competition among operators;
- **Badges:** a form of reward that can show performance, status, and increase motivation;
- **Challenges:** a way to give direction and a sense of purpose to the players.

2.3.2. Types of Motivation

As mentioned earlier, gamification can be used as a tool to increase motivation. Zichermann and Cunningham [49] divide motivation into two components: intrinsic and extrinsic motivation. Intrinsic motivation refers to the personal goals of a person (self-motivation), and does not depend on external factors, while extrinsic motivation is directly determined by external factors, such as incentives. According to Zichermann and Cunningham [49], gamification mostly focuses on the enhancement of extrinsic motivation, which is most effective when aligned with the intrinsic motivation of an individual.

Gamification could also potentially negatively affect motivation, for instance when a leaderboard is used. According to Vansteenkiste and Deci [52], the “winners” (players at the top of the leaderboard) are more motivated than the “losers” (lowest ranks of the leaderboard). However, positive feedback can alleviate the negative effects of losing, and can compensate for lacking intrinsic motivation through increasing extrinsic motivation.

2.3.3. Ethical Implications of Gamification

Due to the nature of gamification, the concept poses several problems, as certain tactics used to implement it are considered unethical. Kim and Werbach [53] describe four most prevalent ethical issues with gamification mentioned in literature. The first is that gamification takes unfair advantage of workers. For instance, in the previously mentioned example of gamification implementation, where a virus was studied with the help of people playing a mobile game [51], the work of the participants is under-recognised, as they are mentioned as a group in the scientific publication, without individual recognition. Kim and Werbach [53] also state that gamified management techniques are exploitative, as the incentives given by companies are fictional.

The second ethical problem with gamification is the possibility that its implementation is manipulative. Kim and Werbach [53] ask the question of whether it is ethical to force a worker to be part of a gamified environment, when one important aspect of playing a game is voluntarily choosing to do so (autonomy). It is therefore required that gamified applications are transparent towards the people encompassed by them: all parties should be aware of the implications of being part of such a work environment.

Another issue mentioned by Kim and Werbach [53] is that a gamified environment can unintentionally produce physiological harm to participating subjects, leaderboards being given as an example. These can introduce anxiety and job security fears into workers, and competition can make the work environment stressful. Thus, the social context and interactions between game elements and the work environment have to be considered and analysed in order to avoid unintended harm on the workers.

Finally, the literature surveyed by Kim and Werbach [53] raises the ethical consideration of the negative effect that gamification has on the morality of workers and their character. A reward-based system can change the values of people, therefore, the authors of the paper recommend caution when implementing gamification in “serious” environments such that indifference towards human values is not cultivated. Furthermore, operators might downplay risk when dealing with fictional threats, which might prove problematic when required to suddenly shift to a different risk management strategy in case of automation failure.

2.4. Principles of Display Design

The literature study at hand is focused around developing a tool that would enhance engagement and situational awareness in air traffic controllers that supervise automation by using the concepts of threat image projection and gamification. In order to achieve this, documentation is needed on the design of displays in an ATC context, and how to best achieve the intended effects. One fundamental work on display design that should be considered are the thirteen principles of display design by Wickens et al. [54], described below with examples within air traffic control.

1. **Make displays legible/audible;** this implies the transmission of information in a clear manner, such as aircraft identification number (ACID), velocity etc.
2. **Avoid absolute judgement limits;** binary signals signals (yes/no) instead of several levels that could cause confusion. STCA is a good example of this, where the are two or three discrete states implemented: no conflict within 2 minutes, conflict within two minutes, and conflict within one minute.
3. **Top-down processing;** people will interpret signals according to their expectations, thus unexpected signals might be wrongly perceived. For example, the colour of elements in a display needs to match the operator expectations (red for conflicts).
4. **Redundancy gain;** repeated information (under multiple forms) is better perceived. For example, fictional aircraft can be distinct in both shape and colour from real aircraft.
5. **Discriminability: Similarity causes confusion;** make signals easily distinguishable. This is relevant to the research at hand as fictional aircraft must be diferenciated from real traffic.
6. **Principle of pictorial realism;** display should look and behave as expected (RADAR screen, aircraft movement).
7. **Principle of the moving part;** the moving elements of the display should match operator expectations. For example, in an air traffic control environment, choosing a reference frame in which aircraft move while waypoints are fixed is expected. On the other hand, a pilot expects to see waypoints ”moving” on the screen while the location of the aircraft remains in the centre.
8. **Minimise information access cost;** information should be readily available and easy to access. In the context of ATC, the aircraft labels should be designed such that the operator can obtain most information through a brief glance.
9. **Proximity compatibility principle;** similar information should be grouped together, as humans are good at pattern recognition. This can apply to aircraft labels, where unexpected behaviour can be noticed if the label displays different information from other aircraft (for example, low velocity).

10. **Principle of multiple resources;** information can come from several sources, i.e. both auditory and visual information. The STCA system alerts the operator both visually and audibly.

11. **Principle of knowledge in the world;** provide the user with all possible options at all times. This principle relates to ecological interface design, meaning that the operator should be aware of their options for controlling aircraft: heading, speed, altitude command options and ranges portrayed clearly.

12. **Principle of predictive aiding;** humans should be help with predicting future outcomes. For example, showing the previous position of the aircraft can help operators predict future behaviour, such as turning radius.

13. **Principle of consistency;** displays should be consistent in their design, for example all aircraft should be controlled in the same way.

On top of these general principles, Wickens et al. [55] also mention guidelines for monitoring display design:

1. **Analogue versus digital:** if a variable is continuously changing, an analogue indicator might be more appropriate and easier to supervise;

2. **Analogue form and direction:** analogue displays should be designed carefully using the thirteen principles, with focus on principle no. 3 (display should portray information in an expected way), no. 6 (display should look as expected), and no. 7 (moving part should be expected);

3. **Prediction and sluggishness:** slowly changing variables can better be predicted; displays that support prediction are desirable.

2.5. Ecological Interface Design

According to Vicente and Rasmussen [56], ecological interface design (EID) represents a set of design principles that aim to create displays that support and enhance human cognition. This is achieved by analysing the control task through the lens of the behaviour categorisation framework described by Rasmussen [57]. This divides cognitive behaviour in three categories, briefly explained below:

- **Skill based behaviour:** based on reflexes and memory, it implies non-conscious actions and control;

- **Rule based behaviour:** implies the execution of actions in pre-defined situations (i.e., following procedures);

- **Knowledge based behaviour:** this type of behaviour is used in unfamiliar situations and requires the highest degree of workload.

While the focus of the research at hand is not the design of a gamified ecological interface, using the the design principles presented within the EID framework can ensure an implementation of gamification within ATC that does not hinder controller cognitive abilities. Future design iterations could focus more on the relationship between EID and gamification to obtain the combined benefits of both concepts.

2.6. Main Conclusions of Literature Survey

Most literature agrees that a high level of automation in a work environment has a negative impact on the cognitive abilities of humans. The lack of engagement and the presence of boredom induces situational awareness and vigilance issues, which translate to a lower supervisory control performance. Moreover, the lack of practice that a monitoring role entails produces skill degradation.

A potential solution to this problem could be the concept of gamification: the use of game elements in non-game related applications. This concept is currently being used in airport security screening with the scope of maintaining operator vigilance by planting fictional threats within the luggage flow. Operators must identify the fictional threats, and are evaluated or rewarded based on their performance. The concepts have great potential of providing a solution for the air traffic control environment.

The current air traffic displays therefore need to be adapted. In order to produce a competent display, the thirteen design principles need to be taken into consideration. Furthermore, as the display would be a monitoring display, it needs to account for the factors that govern a supervisory task: legibility, information presentation methods and prediction aiding. Ecological interface design methods can also be used to analyse the problem and produce a solution that does not hinder

3

Proposed Concept and Design

The following section presents the proposed concept tool resulted from the use of implementing gamification within a highly automatised air traffic control work environment.

3.1. Preliminary Concept

As previously stated, the concept of threat image projection entails the use of fictional threats in order to maintain engagement among operators. The main threat-like events that occur in air traffic control are aircraft conflicts, meaning two or more aircraft that are on a course that would lead to the loss of separation (less than five nautical miles between them or less than 1000 ft vertical separation).

Therefore, the purposed concept for an engagement, vigilance and situation awareness tool consists of the introduction of ”fictional aircraft” within the automated work environment of an air traffic controller. These fictional aircraft, shown along with the real traffic, need to be routed by controllers through the sector as if they were real: maintaining separation from the other aircraft and routing them to their exit waypoint.

Fictional aircraft would appear different from real traffic as to help controllers maintain the efficient supervision of actual traffic, as seen in Figure 3.1. It should be noted that this is an initial design starting point concept and does not portray the final design of the tool. The design of the concept tool presented in the sections that follow is made within the environment of air traffic control simulation software developed within the Control and Simulation department of the TU Delft Faculty of Aerospace Engineering.

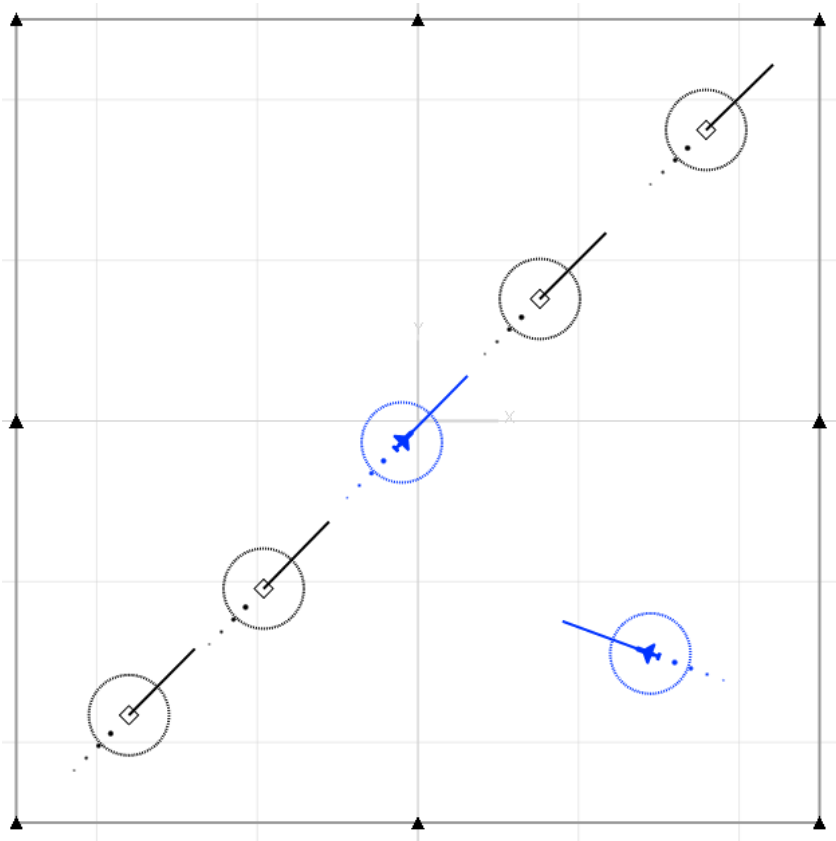


Figure 3.1: Initial concept visualisation for the implementation of threat image projection within an air traffic control environment. The fictional aircraft are differentiated from real aircraft through their colour: fictional aircraft are blue. The circles around the aircraft have a radius of 2.5 NM. Note that this figure portrays a preliminary concept and does not represent the final design.

3.2. Control Task Analysis

The air traffic control work environment is highly complex and controlled, as the task performed within is critical for the safety of the whole aviation industry. Introducing a new tool within this work space could potentially disrupt information flow and the control tasks, therefore an analysis of the work environment must be performed.

The proposed tool would be implemented within a highly automated work environment, in which an automatic conflict detection and resolution system would handle all aircraft in an airspace sector. The automation would act according to the information flow diagram presented in Figure 3.2. The conflict detection system uses data from several sources, such as ADS-B or RADAR, to determine whether two or more aircraft are in conflict. If this is the case, the automated system uses priority rules, sector geometry and optimisation algorithms to create a course of action which it then implements without needing the approval of the system supervisor.

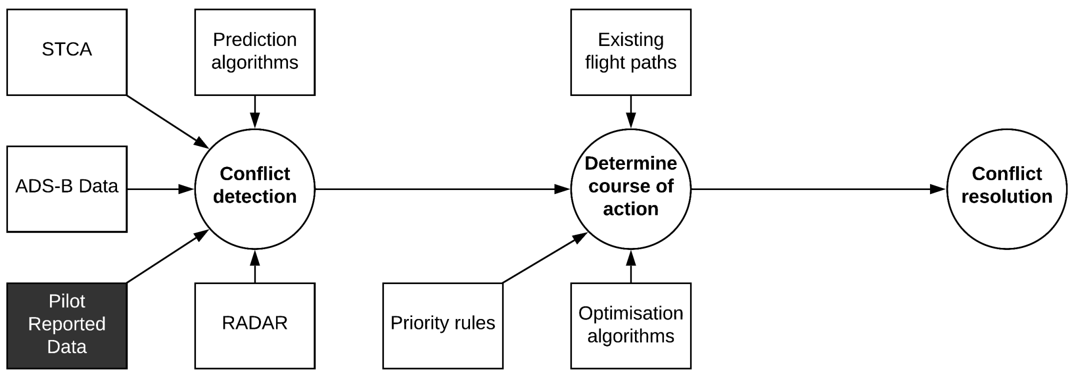


Figure 3.2: Information flow diagram for an automatic conflict detection system. Dark grey boxes represent human tasks and white boxes represent automated tasks or information gathering.

The role of an air traffic controller would also be different than present times, as a transition from an active to a passive role fundamentally changes the control task. This can be seen in the information flow diagram presented in Figure 3.3. The air traffic controller would use information from the same sources as the automated system to monitor automation. In case an automation failure is detected, then the operator must take over manual control in order to limit any consequences of the failure.

From the cognitive point of view, the transition from supervisory to manual control represents a critical moment for an operator, as their task and strategy need to change. This can produce a decrease in manual control performance immediately after the operator intervenes, especially if, prior to the intervention, situation awareness and vigilance are lower due to the monotonous nature of supervisory control in a highly automated environment.

As mentioned in section 2.1.3, research is performed within SESAR that seeks to explore and mitigate the effects of a high level of automation on the cognitive abilities of the supervisor. It is therefore known that maintaining situation awareness and vigilance is ben-

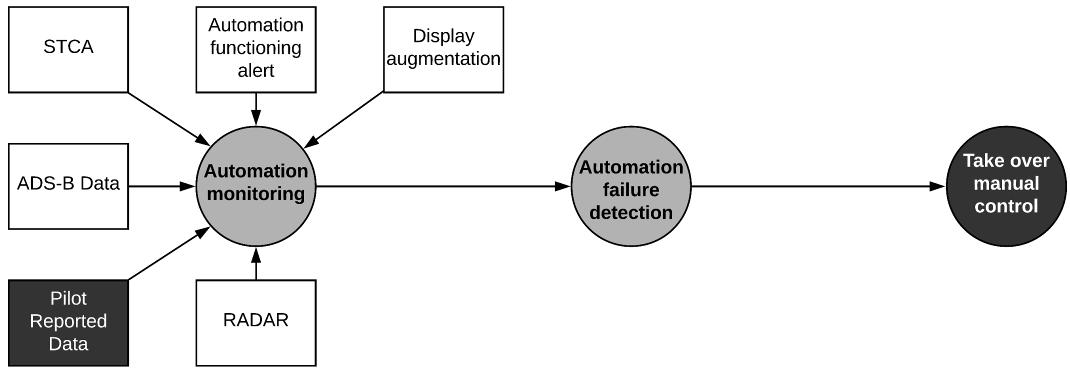


Figure 3.3: Information flow diagram for a system supervisor for detection of exceptional events. Dark grey boxes represent human tasks, light grey boxes are tasks performed by both automation and humans, and white boxes represent automated tasks or information gathering.

eficial for spotting mishaps in automation. Therefore, on top of the hypothesised benefits on cognitive abilities during automation supervision, the proposed concept could also improve manual performance immediately after control takeover. This effect is described in Figure 3.4.

The decision ladder diagrams are based on the ones presented by Borst et al. in [58] for decision support tools. They incorporate the type of behaviour employed by operators when performing decisions. In the supervisory control case, a decision process is activated if the operator believes a mishap has been detected (skill-based behaviour). The operator then uses rule-based and knowledge-based behaviour to determine if intervention is necessary by assessing the potential options and consequences.

If the situation awareness and vigilance of the operator are low, a transition to manual control would determine the initiation of a new decision process represented in Figure 3.4 by a separate decision ladder diagram. The operator would have to begin solving conflicts, and would have to gain knowledge of the situation by navigating the initial steps after activation. However, if situation awareness is high, then the first steps after activation could potentially be skipped, thus creating a shortcut that would lessen decision time and mental workload, leading to improved manual control performance.

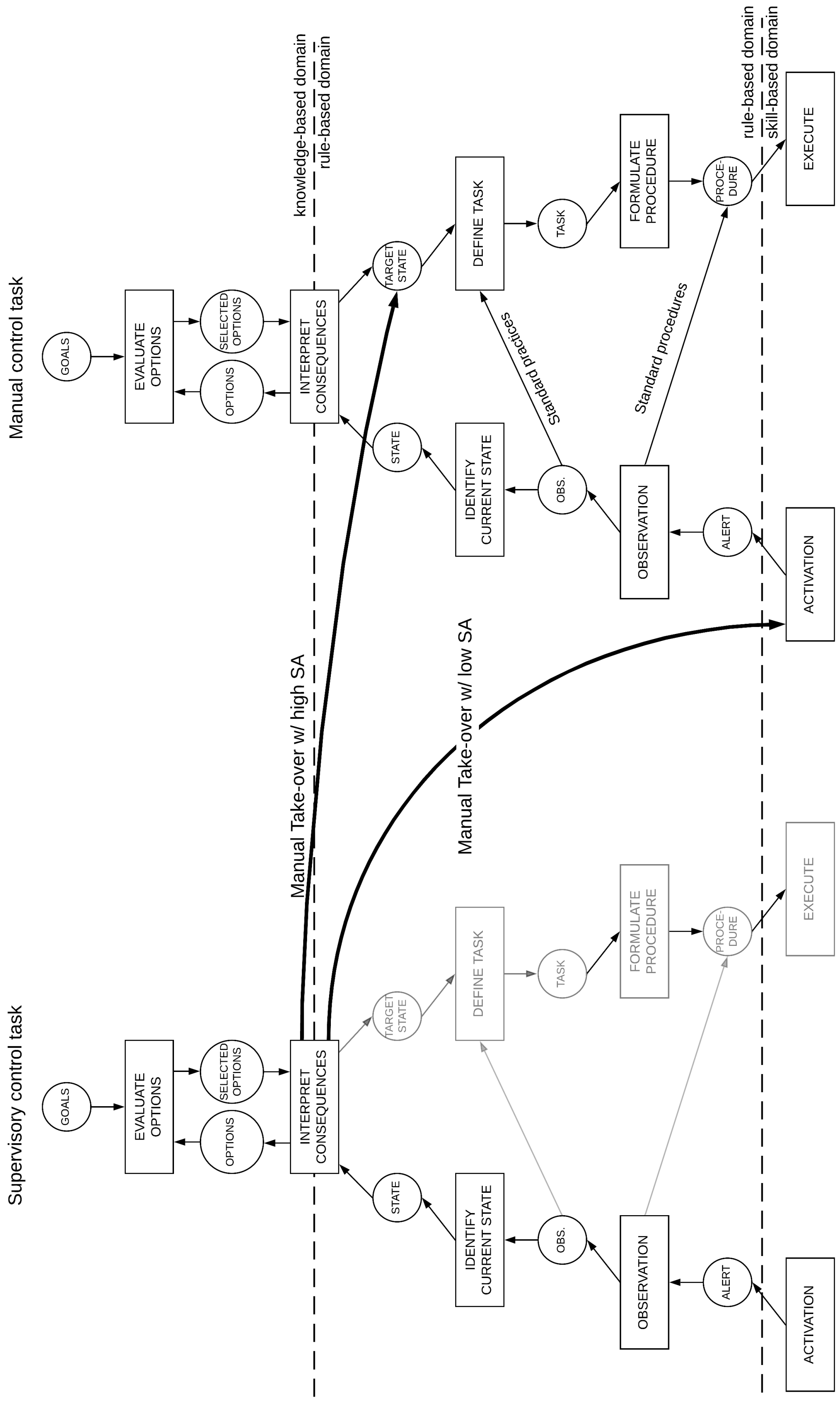


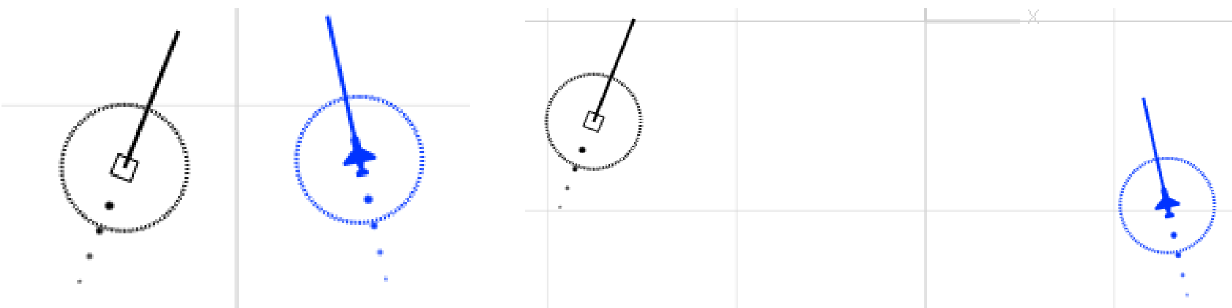
Figure 3.4: Control task analysis using decision ladders portraying the indented effect of the implementation of the proposed tool in a supervisory control ATC environment.

3.3. Threat Image Projection in ATC

As explained in section 2.2, there are three parameters that govern the implementation of threat image projection in baggage screening: view difficulty, superposition levels and view complexity. These directly influence the single-threat workload experienced by the operator when performing supervisory control.

These factors could be transposed within the air traffic control work space in the following way:

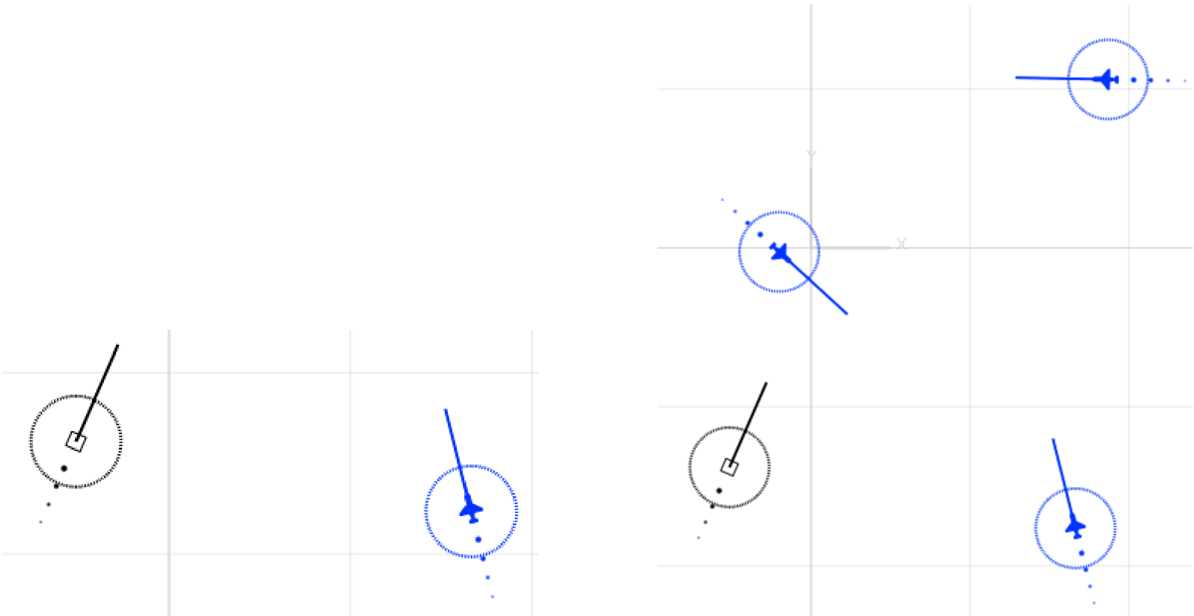
- **View difficulty:** the situation portrayed in Figure 2.3 can be compared to how noticeable the fictional conflicts are. For example, if fictional aircraft (blue) are close to real aircraft (Figure 3.5a), then the conflicts are more noticeable and draw the attention of the operator. However, if fictional aircraft are far from the real traffic (Figure 3.5b), the future conflict is more difficult to detect at the portrayed moment.
- **Superposition levels:** The superposition location of the threat on the scanned baggage can alter the difficulty level, as seen in Figure 2.4. This can be compared to a situation in which a conflict between a real and fictional aircraft is hidden in plain sight by the presence of other aircraft in the sector. An example of this is portrayed in Figure 3.6: the conflicting aircraft are easy to spot in Figure 3.6a, but more difficult to determine in Figure 3.6b, even though only two aircraft will produce a conflict in the latter.
- **View complexity:** The density of objects in a bag produces a higher workload for the screening operator, as more objects need to be analysed, as seen in Figure 2.5. This can be compared to the amount of aircraft present in the sector at one time. Figure 3.7a portrays a situation where the traffic density is small, and the amount of fictional aircraft as well, and Figure 3.7b depicts a more dense traffic situation with an increased number of fictional aircraft.



(a) (b)

Figure 3.5: View difficulty factor of TIP applied to ATC.

The last factor that influences the workload of an operator is the threat frequency. In terms of air traffic control, this can be implemented as the number of fictional aircraft present at one time in the sector: more fictional aircraft result in more fictional conflicts and threats.



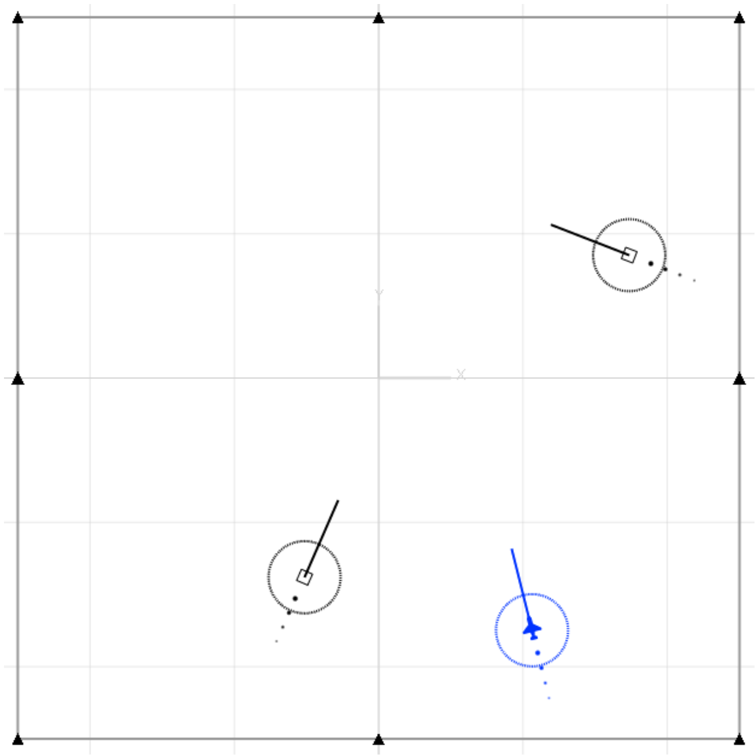
(a) (b)

Figure 3.6: Superposition factor of TIP applied to ATC.

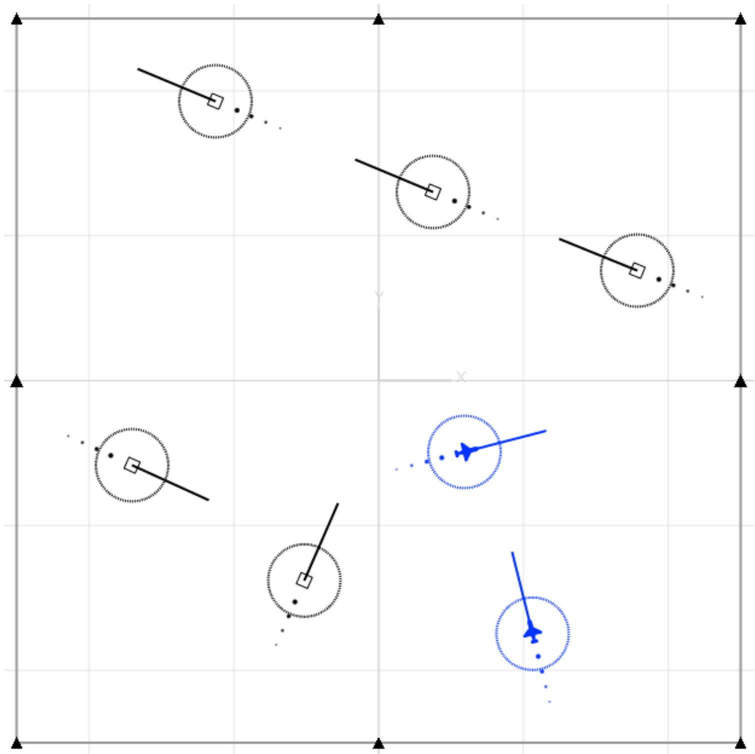
There is one major difference in the proposed implementation of threat image projection in air traffic control as opposed to the implementation in baggage screening. In the latter, operators are not aware that the threat is fictional until they have resolved it. In the proposed implementation for ATC, operators would be fully aware that the threat is fictional, and that the primary goal is to prevent conflicts in real traffic.

This could mean that the observed benefits of threat image projection might not translate directly to ATC as the implementation is different. From an ethical point of view, the proposed implementation might be preferred as operators are fully aware of the goals of the aid tool and are not manipulated into thinking the threats are real. This could also help with the stress induced by threat image projection in airport security personnel.

On the other hand, the motivation for resolving fictional threats might have to come from a different source in case of ATC, as in the luggage screening application, operators are motivated by the fact that any threat has a small probability of being real. In the ATC case, operators might have to be motivated through other means (a reward, penalty based system or internal motivation), otherwise their acceptance of the fictional aircraft aid might be low.



(a)



(b)

Figure 3.7: View complexity factor of TIP applied to ATC.

3.4. Implementation of Fictional Aircraft

The following section presents the design process and decisions made for the implementation of fictional aircraft within an air traffic control environment. In order to facilitate the conduction of an experiment, the SectorX ATC simulator developed at TU Delft is used as a simulation platform.

3.4.1. Fictional Aircraft Visual Design Considerations

The design of fictional aircraft is based on the implementation of aircraft within the SectorX software. The proposed tool design is developed based on the expectation that the experiment is performed using students (explained in Chapter 4), and therefore the interface is designed such that it better suits operators with low experience in air traffic control.

The aircraft in SectorX are portrayed as shown in Figure 3.8. The current position is marked intuitively by an aircraft shaped icon, while the aircraft heading and velocity is represented by a line. The horizontal separation criteria of 5 NM is shown as a dotted circle around the aircraft position with a radius of 2.5NM (the circles should not intersect). In accordance with the principle of predictive aiding, the previous positions of the aircraft are presented as trailing points.

An aircraft label is attached to the current position, and presents information in a clear and direct way to the operator. The first line represents the flight number, and the second line portrays the current altitude (flight level) and velocity (knots) of the aircraft. The last line shows the target exit waypoint and the aircraft size (S for small, M for medium, L for large).

This method of portraying aircraft is similar to actual air traffic displays, as shown in Figure 3.9. Here, aircraft are represented as more simple shapes (such as a square or circle), and with no separation requirement perimeter. The latter is an important element when considering low experience operators (students), and is therefore kept enabled within the present iteration of the ATC aid tool.

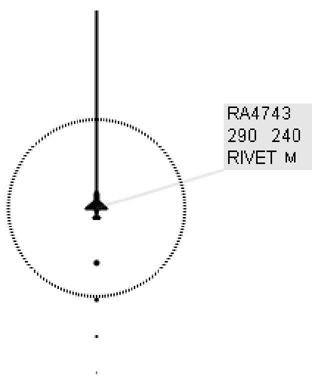


Figure 3.8: Aircraft blimp as implemented in SectorX.

Figure 3.9: Typical air traffic control display used within the industry: ATC software developed by INDRA [59].

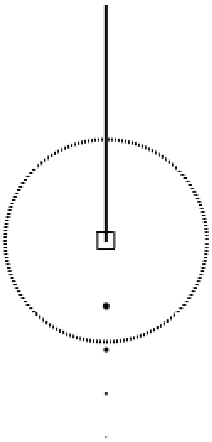


Figure 3.10: Real aircraft display element within the simulation environment of SectorX.

However, if considering the display design principles of discriminability and redundancy gain, and considering that fictional aircraft should be clearly distinguishable from real aircraft, it was decided to differentiate the two types in both form and colour.

3.4.2. Aircraft Element Shape

Differentiating the two types of aircraft display elements implies changing the icon for one, or both within the SectorX simulator. In order to increase the fidelity of the simulator, as well as comply with the principle of pictorial realism (i.e., the display should be similar to the ones used in upper area traffic control), real aircraft will be depicted as a hollow square, similar to their portrayal within the INDRA work environment (Figure 3.9), as shown in Figure 3.10.

It was also decided to keep the original aircraft-like icon for fictional aircraft (Figure 3.8). This encourages the idea that these elements should be regarded as actual aircraft, while also

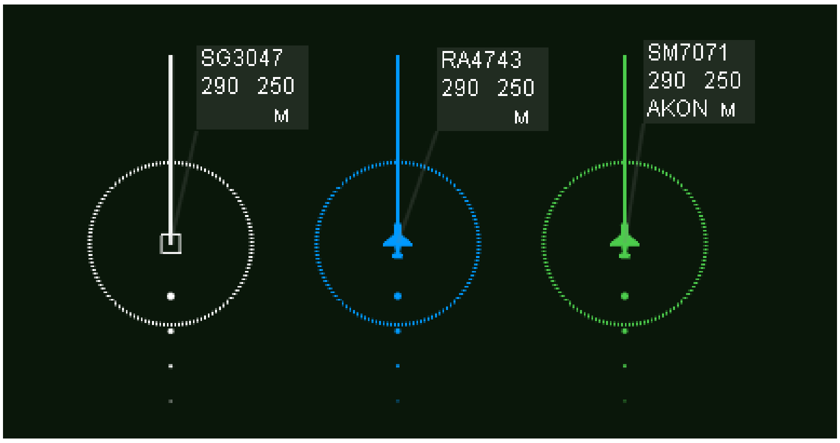


Figure 3.11: Colour design for aircraft elements within the SectorX simulation environment. From the left: real aircraft, fictional aircraft, fictional aircraft on-course.

producing a noticeable visual difference. In order to aid less experienced controllers during the experimental phase, the velocity vector and the horizontal separation circle are kept as part of the display element in both fictional and real aircraft representations.

3.4.3. Aircraft Element Colour

Colour is an important design consideration, related to the display design principles of legibility and discriminability. According to Ahlstrom [60], colour can influence human factors such as salience recognition, which in turn can affect workload and reaction time in an air traffic control environment. The research concludes on a certain pallet of colours that are highly noticeable and improve human performance. Considering a dark background, white and bright variations of green and blue are the most luminescent colours.

Within the SectorX simulation environment, green is already used to inform operators that an aircraft is on course to its exit waypoint. This is not a feature within an actual ATC display, as it would be difficult to determine when an aircraft is on course given the variety of flight paths and waypoints. Therefore, this feature is deactivated for real aircraft.

However, this feature is kept for fictional aircraft for the following reasons: a target exit waypoint can be set and ”on-course” indication can be implemented, and it helps with not diverting more attention than required from the main task of supervising real traffic. Therefore, the remaining two colours that Ahlstrom [60] recommends are white and bright blue.

It was therefore decided to use white to colour the real aircraft icons, as it is common practice in the ATC domain, and is therefore according to expectation. Fictional aircraft are portrayed using a variation of the light blue colour. The final aircraft element design is presented in Figure 3.11, with the RGB values of the selected colours presented in Table 3.1.

The final consideration relating to colour use in aircraft elements is the colours used by STCA to warn operators of imminent conflicts. It is an important tool when supervising

Table 3.1: RGB colour values for real aircraft and fictional aircraft in on/off course situations.

Display element	RGB values		
Real aircraft	0	0	0
Fictional aircraft (off course)	0	153	255
Fictional aircraft (on course)	77	204	77

automation, as it can draw the attention of an operator to a potential conflict or automation failure. Furthermore, according to the display design principle of predictive aiding, it is an important tool for aiding controllers predict future situations.

In the context of the implementation of fictional aircraft, the implementation of STCA within SectorX needs to be modified. Initial behaviour consists of changing the colour of a pair of any two aircraft (regardless of type) if a conflict is imminent (orange or red if loss of separation occurs within two and one minute respectively). However, to avoid confusion, fictional aircraft should not determine a STCA response in real aircraft, as this could lead to false information transmission (real traffic is in conflict when it is not). Therefore, one option could be that, for conflicts between fictional and real aircraft, only the fictional aircraft experience a change in colour.

As for actual traffic, there are two options: enable STCA for conflicts for real aircraft pairs only or disable STCA entirely. The first option could be the most beneficial in terms of the implementation of the gamified tool in an ATC environment, although more research is required on the influence of the two tools on each other. In a controlled experimental environment, in order to mitigate the effect of this influence, STCA is deactivated entirely for both types of aircraft. However, for an actual implementation, it is recommended that STCA is implemented through the methods described above.

3.5. Other Design Considerations

The following section presents the remaining design considerations taken into account. These include the behaviour of the aid tool in case automation experiences a failure or manual control is demanded, as well as the number and pattern of fictional aircraft within the air sector.

3.5.1. Automation Failure Behaviour

The purpose of the aid tool is to enhance and maintain situation awareness in air traffic controllers performing a supervisory role. As there is always a possibility for automation failure and a requirement for manual intervention, there needs to be a procedure for dealing with fictional aircraft in this situation.

There could potentially be two ways through which manual control is given to the operator: an automatic mode, through which failure is detected automatically and the operator is requested to intervene, and a manual mode, in which the operator decides to intervene without

Table 3.2: Hypothesized effect of fictional aircraft pattern and density on situation awareness, workload, and control performance compared to the unaided supervisory control case.

	Crossing		**Among**		**Combined**	
Low density	SA	↑	SA	=	SA	⇈
	Workload	↑	Workload	=	Workload	↑
	Performance	↑	Performance	=	Performance	⇈
High density	SA	⇊	SA	↑	SA	↓
	Workload	⇈	Workload	↑	Workload	↑
	Performance	⇊	Performance	↑	Performance	↓

being prompt to. In both of these cases, fictional aircraft disappear from the radar screen in order to permit the focus of the operator on manually controlling real traffic. This is explained and portrayed in Appendix A, as well as a conceptual representation of the interface of the proposed tool.

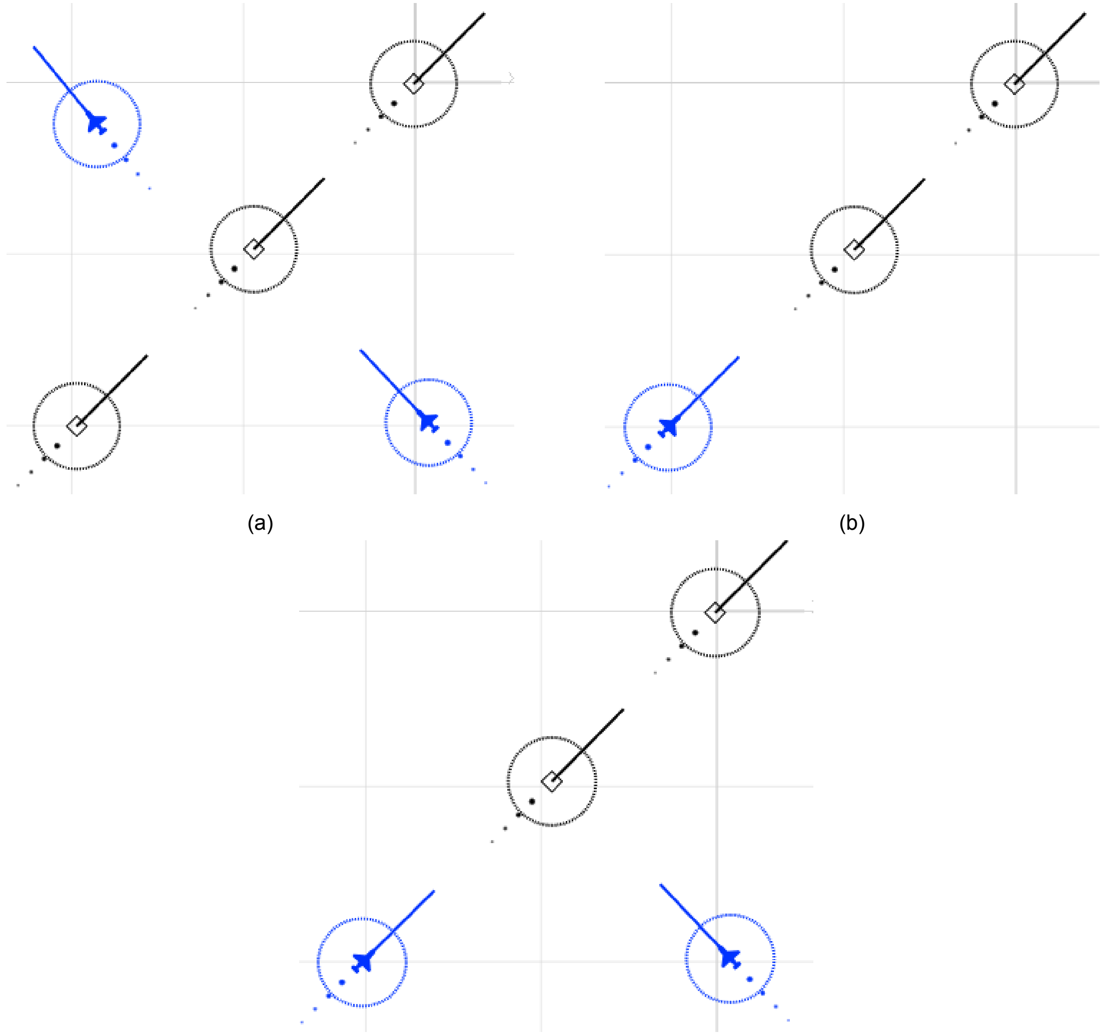
3.5.2. Fictional Aircraft Density and Pattern

Another important design aspect is the density of fictional aircraft within the air sector. Workload and supervisory control performance can easily be influenced by this, as a low density can lead to lower than ideal workload conditions and therefore decreased engagement. On the other hand, a high fictional aircraft density can lead to high supervisory and manual control workload and a lack of focus on real aircraft traffic. Due to the importance of this factor, it is decided that it will be tested experimentally: subjects will use the aid tool with varying density of fictional aircraft. This could provide information on whether the predicted effects of the number of aircraft on workload are reasonable.

Fictional aircraft pattern is another important design aspect, as it can influence cognitive factors such as workload, situation awareness and cognitive tunnelling. Upper airspace is generally organised using airways to which most commercial flights are assigned. Depending on the air sector, this can range from a couple of main routes to several routes.

In general, there are three fictional aircraft patterns that can be applied to the simulation tool: fictional aircraft cross real traffic (Figure 3.12a), fictional aircraft are among real traffic (Figure 3.12b), and a combination of the two (Figure 3.12c). There are several advantages and disadvantages for each, especially when considering the fictional aircraft density, summarised in Table 3.2.

Compared to a baseline situation in which an operator must supervise traffic unaided (no fictional aircraft), it is predicted that the overall workload will generally increase in all situations, except in the low density, among pattern, as conflicts will be more rare. Furthermore, the presence of a high number of fictional aircraft is predicted to have negative effects on situation awareness and performance, as this would lead to a large amount of conflicts that require the attention of the operator, thus preventing them from focusing on real traffic.



(a) (b)

(c)

Figure 3.12: Fictional aircraft patterns: crossing (a), among (b) and combined (c).

The best pattern to be considered is most probably the combined pattern, as fictional aircraft are present in all areas of the screen, thus mitigating the effect of attention tunnelling. At the same fictional aircraft density, the combined pattern would perform better than the crossing and among patterns as the extra workload is not excessively high while still maintaining operator engagement. It will therefore be used by the gamified aid tool for the distribution of fictional aircraft.

3.5.3. Operator Motivation

As mentioned previously, one major difference between the implementation of TIP in baggage screening and the proposal presented in this report is that, in the latter, system supervisors are aware of the fictional nature of certain aircraft on the screen. This was done for ethical reasons, as the design at hand seeks to implement gamification in a non-manipulative and transparent manner. However, this could affect the motivation for controlling these aircraft, as there is no apparent goal in doing so.

Literature presented in Chapter 2 concludes that positive motivation is more effective than motivation through penalties. Therefore, one potential solution to this problem could be another element of gamification: scores. Operators will be shown a live score equivalent to their control performance as direct feedback. This scoring system is already implemented within SectorX and is a function of several variables such as the instances of losses of separation or whether the aircraft are on trajectory or not. This could be complemented by rewarding operators that achieve a high average score over a certain period of time. Scores are also shown to enhance internal motivation as operators seek to continuously improve or maintain a good level of performance.

4

Experimental Design

As the implementation of threat image projection within ATC is novel to this field, an experiment must be performed in order to test whether it is beneficial and achieves the intended effects. The following chapter presents the hypotheses derived from the research questions as well as the experiment development process and results.

4.1. Hypotheses

Before the experimental setup is developed, the hypotheses that need to be tested must be stated. These are produced in relation with the sub-questions that are to be answered experimentally (SQ-7, SQ-8 and SQ-9).

HP-1 The implementation of fictional aircraft within an ATC supervisory control environment improves the supervisory control performance of operators.

HP-2 The implementation of fictional aircraft within an ATC supervisory control environment improves manual control performance in case of automation failure.

HP-3 The implementation of fictional aircraft within an ATC supervisory control environment improves situational awareness in operators.

HP-4 The implementation of a high number of fictional aircraft will produce attention tunnelling and therefore will worsen performance.

HP-5 The implementation of fictional aircraft within an ATC supervisory control environment partially mitigates the effects of skill degradation.

4.2. Considered Experiment Setups

Over the course of the literature study and experiment design, there were several options that were considered for testing the hypotheses. Initially, the idea of including the effect of alerting the operator of a potential failure was considered. This implied that there would be two possible situations: the system detects automation failure and alerts the operator, and the system does not detect failure and thus fails to alert the system supervisor (i.e., automation fails silently).

As no system is without the possibility for failure, this experiment could have covered the worst-case scenario in which operator intervention is required. However, as the proposed concept is in the early research stages, it was decided to focus the experiment on the potential of gamification for improving ATC supervisory control performance. Furthermore, the limited access to experiment subjects determined an effort to concentrate on the factors that most likely influence the design of the fictional aircraft tool, such as the density of aircraft in the sector.

Furthermore, asking subjects to find a conflict and then take over manually would have had an influence on the controlability of the experiment, as, depending on the time it would take subjects to find the conflict, the situation might have developed into a more difficult one. This would mean that the subjects that took a longer time to find the conflict would most probably perform worse during the manual control phase as well. As it is difficult to equalise the training and skill level among participants before the experiment, this effect would produce a large variation in the results.

4.3. Final Experiment Format

The experimental version of the tool is different than the envisioned design for an air traffic control environment, as it was decided that the experiment should minimise the influence of other factors on the measured data. The following section presents the finalised experiment format and setup, as well as the decision process that was used to reach them.

4.3.1. Experiment Participants and Matrix

The overall experimental setup is highly dependent on the number of participants that will be involved. A larger sample size leads to higher quality statistical results, however, at the time of the writing of this report, it is uncertain whether a large enough subject pool is accessible. Therefore, two experiment matrices were developed in order to account for a both possibilities.

It should be noted that the experiment setup follows a mixed design, in which results are compared both within and between participants. There are three cases to be tested:

Case 1 **Baseline:** the control case which represents unaided supervisory control: the test subject must supervise and control air traffic within a sector without the presence of external tools;

Case 2 **Low number of fictional aircraft:** the test subject must supervise and control air traffic as well as control a low number of fictional aircraft within a sector;

Case 3 **High number of fictional aircraft:** the test subject must supervise and control air traffic as well as control a high number of fictional aircraft within a sector;

Using these cases, the experiment matrix presented in Table 4.1 was developed. A major consideration was the learning effect that participants would experience, in which case the second run would always display better performance. In order to counteract this, the participants will be divided into two major groups: a half that does the baseline case first, and a half that does the fictional aircraft scenario first. The within subject part of the experiment compares the relative performance difference of each subject between the baseline and the fictional aircraft cases. It should also be noted that the matrix presented in Table 4.1 is a single cell and thus is recurring in function of the total number of participants.

Table 4.1: Single cell of the proposed experiment matrix in case a large number of subjects is available.

Subject No.	**Run #1**	**Run #2**
1	Baseline	Low # fictional a/c
2	Baseline	High # fictional a/c
3	Low # fictional a/c	Baseline
4	High # fictional a/c	Baseline
	⋮	

As explained later, the length of the experiment per subject was also taken into consideration when developing the experiment setup. One run is supposed to last approximately 70 minutes, in which case, more than two runs would constrain both the willingness to participate and the availability of test subjects. It was therefore decided to limit the number of runs to two per participant.

In case a large number of participants is not available, this experiment matrix cannot be used, as there will not be enough data points per group. In this case, the matrix has to be limited, and therefore only two conditions are defined: subjects that perform the baseline and the low number of fictional aircraft cases, and subjects that perform the baseline and the high number of fictional aircraft cases, as presented in Table 4.2.

Table 4.2: Single cell of the proposed experiment matrix in case a small number of subjects is available.

Subject No.	**Run #1**	**Run #2**
1	Baseline	Low # fictional a/c
2	Baseline	High # fictional a/c
	⋮	

This could lead to bias in the results due to the learning effect, especially when the subjects will probably be selected among master students of the TU Delft Faculty of Aerospace Engineering, which have low experience with air traffic control. This could be mitigated through rigorous training or an entrance test (described in a later subsection), as well as the careful selection of participants (students which attended the courses AE4321-15 Air Traffic Management or AE4318 Supervisory Control & Cognitive Systems courses), as prior experience with ATC simulators such as BlueSky[1] or SectorX is desirable.

4.3.2. Simulator Display and Interface

The user interface during the experiment needs to be a controlled environment, as visual elements can easily influence the results. In order to minimise the influence of unexpected effects on subject cognitive performance, some features of SectorX will be disabled, presented below:

- **Advanced prediction aiding**: there are several prediction aids implemented within SectorX, such as a 10 minute future preview. Their implementation within an ATC environment is still being researched, thus they will be deactivated.
- **Short term collision alert:** while STCA is a useful tool and actively used in ATC, it might hinder the isolation of the effect of fictional aircraft. In an experimental setup, STCA influences situation awareness and vigilance, and this effect could be wrongly attributed to the presence of fictional aircraft. Therefore, it is decided that STCA will not be used during the experiment, and the scenarios will be developed such that its absence does not create a difficulty spike.
- **Altitude information and control:** in order to constrain the difficulty of the scenarios, the experiment will be conducted in a 2D environment (top-down view, no altitude dimension).
- **Aircraft velocity control:** as changing the velocity of aircraft during cruise is not generally good practice, and in order to simplify training, aircraft velocity will be kept constant.

The interface during the experiment will be similar to the one presented in Figures 4.1 and 4.2. The most notable difference compared to the concepts presented in Appendix A is that the manual control button is missing. This is because, during the experiment, subjects cannot request manual control at any time, thus achieving better control and standardisation over all the scenarios. Moreover, subjects will be able to report anomalies that occur during a run through the use of the ”Report Anomaly” button, present while automation is functioning (Figure 4.1). It should also be noted that the baseline cases will be run without the presence of gamification elements such as fictional aircraft or score.

4.3.3. Experiment Duration and Phases

As the research project is focused on mitigating loss of situation awareness and vigilance, the length of one experiment is critical to recording data on the potential improvement in perfor-

[1] `https://github.com/TUDelft-CNS-ATM/bluesky`

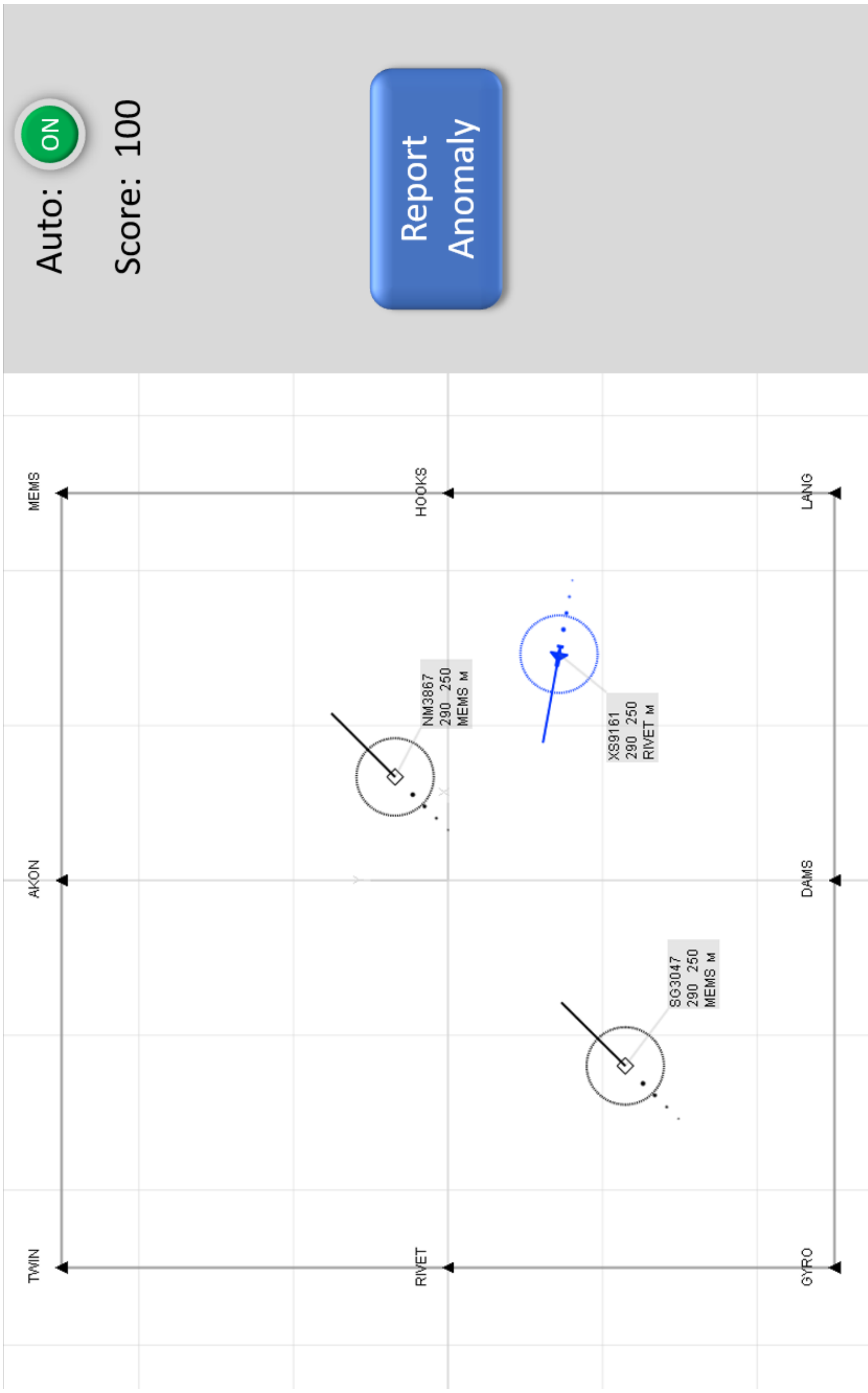


Figure 4.1: Experiment interface while automation is active. Supervisor cannot manually control real aircraft, but can report anomalies that occur during the run.

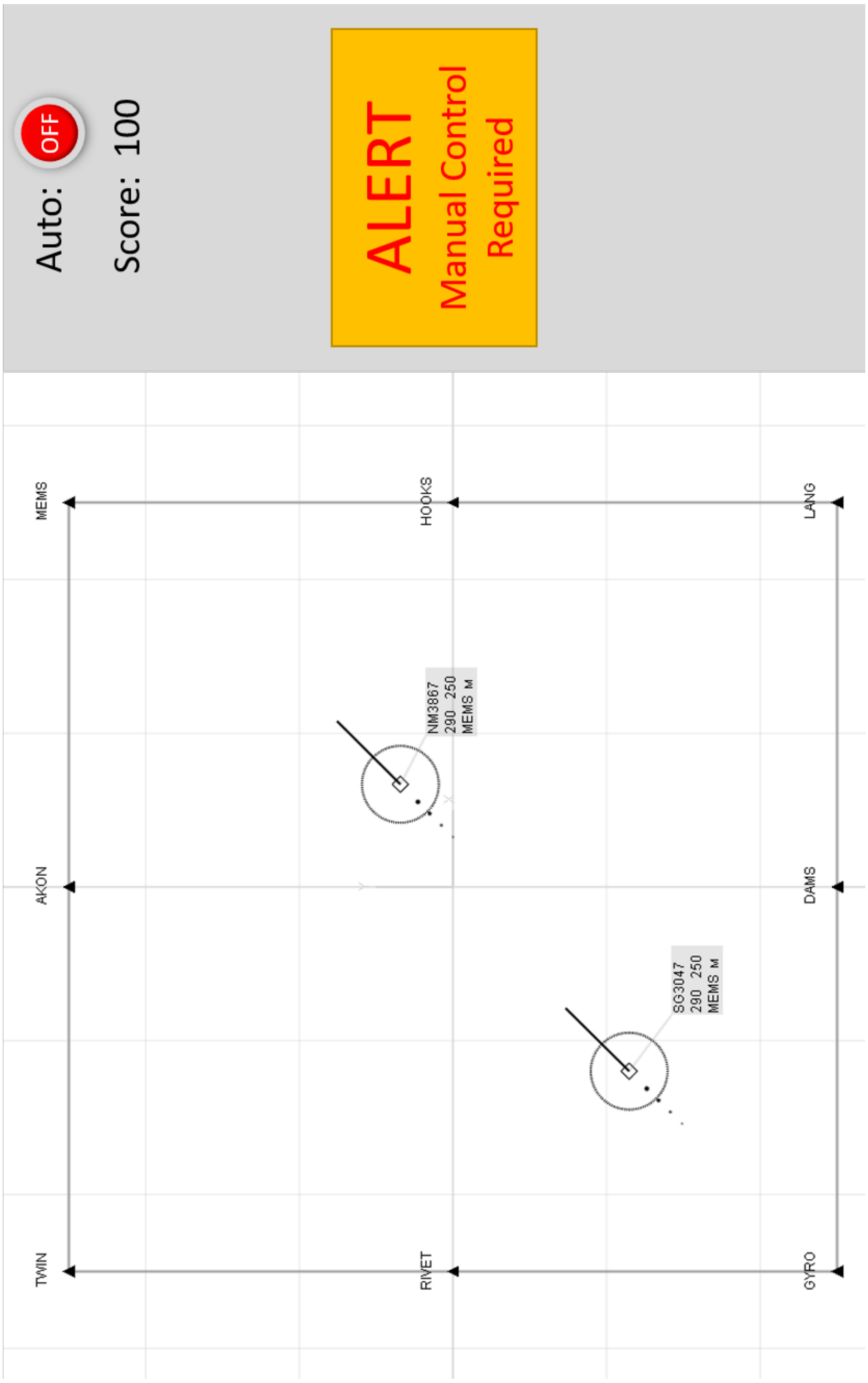


Figure 4.2: Experiment interface when automation failure is simulated. An alert is posted on the screen, subjects can control real aircraft, fictional aircraft are eliminated from the screen.

mance when using gamification in an ATC environment. The baseline case must be designed such that boredom is involved in order to create a meaningful comparison with the cases in which fictional aircraft are involved.

Research on the effect of the length of a monotonous task has been first documented by Mackworth [61] through an experiment involving a monotonous supervision task (supervising a clock). The arm of the clock would occasionally move two steps at a time, and the subjects had to indicate when this happened. The conclusions of this study are presented in Figure 4.3. After the first half an hour, the miss rate was already more than 15%, with the performance experiencing a steep decrease after the first hour.

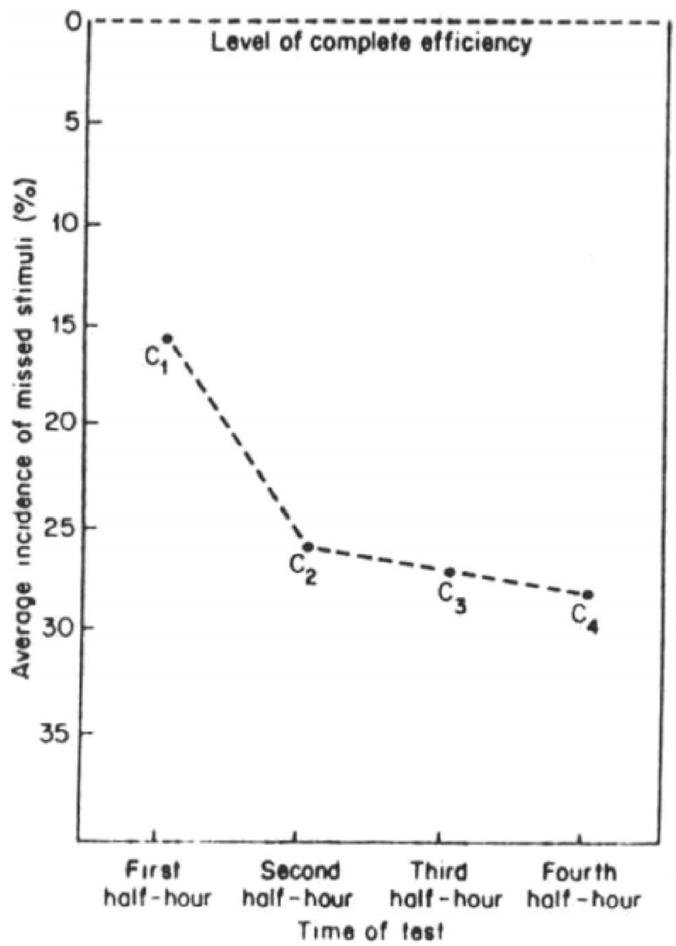


Figure 4.3: Visualisation of loss of performance in function of time during a supervisory control experiment, originally from [61], redacted by Hancock [62].

These conclusions are expanded upon by Hancock [62], which experimentally proves that the design of the task influences the rate of vigilance loss. This is shown in Figure 4.4, where A,B and C represent monotonous tasks with different degrees of engagement. Therefore, a less engaging task (A) can experience a larger decrease in performance level than a better designed task (C).

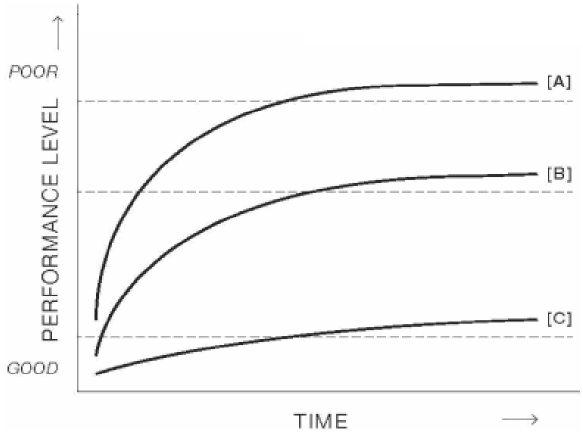


Figure 4.4: Invigilance increment graph as described by Hancock [62]. Performance drops depending on the task performed.

This supports the hypothesis that fictional aircraft, by providing a higher degree of engagement, can lower the effects of boredom in subjects (comparable to case C). Thus, in order to obtain relevant data and a good comparison between cases, the baseline case should be made comparable to case A, as that will show the influence of fictional aircraft most clearly.

An experiment run is therefore divided into three phases: the first phase will consist of a purely supervisory task, in which the test subjects must follow real traffic and control fictional aircraft if these are present. As shown in Figure 4.5, the duration of the first phase will be approximately 15 minutes. The second phase of the run consists of a number of anomalies that will occur on the screen, when the subject should already experience low vigilance. This will allow the measurement of the reaction time of subjects over a period of about 45 minutes, which could be correlated with the loss in supervisory performance in this period. These are better described in the next subsection.

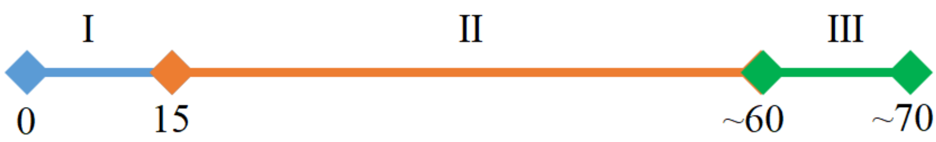


Figure 4.5: Single experiment run timeline divided into parts with timestamps in minutes.

During the first two phases, the subject can only control fictional aircraft if these are present. This is thus hypothesised to induce a considerable decrease in vigilance in the baseline case, which could be mitigated when fictional aircraft are present.

The third, and final phase of a single run is the manual control phase, which is a relatively short period in which the subject is alerted that automation failure occurred and are given control of real aircraft. This phase seeks to observe the behaviour and performance of subjects when faced with a sudden change from supervisory control to manual control. It is hypothesised that fictional aircraft will help lessen the effects of the transition on the cognitive abilities of operators. This phase is kept short, as the transition effects dampen in time as the subjects adapt to the new task.

4.3.4. Automation Implementation and Failure Modes

During the experiment, subjects will have to detect anomalies happening during the second phase. These anomalies consist in real aircraft showing unusual behaviour, or automation giving wrong commands. The subjects will have to report these anomalies through the use of the provided button, which records the time stamp and a short description they provide of the event. This is done to test the reaction time (an indication of situation awareness and vigilance) of operators. These anomalies do not represent a critical automation failure, and thus manual control is not given to the operators when these occur. The types of possible anomalies are presented below.

- **Incorrect exit waypoint:** an aircraft is routed by automation through the incorrect exit waypoint. This anomaly will be more difficult to spot as it will occur at the edge of the screen.
- **Unusual aircraft behaviour:** for example, an aircraft deviates from the route for no apparent reason.
- **Restricted airspace trespass:** SectorX allows the possibility to include restricted airspace within the air sector. An automation anomaly could consist in an aircraft crossing such an airspace.
- **Loss of separation in aircraft:** this could happen at an intersection between two routes, where two aircraft briefly experience loss of separation.

These anomalies can be implemented with varying degrees of conspicuousness. This should be considered when designing the automation, as anomalies that are too easy or difficult to notice will produce unreliable results in terms of reaction time.

As for the implementation of automation within SectorX, there are two options: create an automation algorithm that controls real aircraft, or produce a pseudo-automation through which commands are issued and recorded, and then played back during the experiment. The first method produces more consistent real aircraft movement, while the latter provides increased control over their behaviour. The future stage of the research project will determine which course of action will be taken through preliminary testing and consideration of time constraints. However, the pseudo-automation option is preferred, as it better matches the scope of the research project at hand.

Both methods will have to be implemented by considering standard air traffic controller practices for safely and efficiently routing aircraft through a sector, as automation needs to be predictable. Kim [63] performed a literature study in this scope and suggested several standard practices. The most relevant ones for the application at hand were selected:

- Look-ahead time should be 5-10 minutes;
- High workload conditions require immediate action after conflict detection;
- Low workload conditions imply observing before intervention;
- Safety is the highest priority, especially in high workload conditions;
- Use resolution that requires the least monitoring and coordination;
- Minimise number of commands to aircraft;
- Minimise additional track miles;
- Turn slower aircraft behind for crossing conflicts;
- Solve conflicts on a pair-wise basis;
- Use standard solutions in high workload conditions.

4.3.5. Traffic Density and Pattern

One of the most important aspects of the experiment setup is the traffic scenario design, as they have the potential to influence the cognitive abilities of the controllers to a high degree. In order to produce relevant data, the traffic scenarios need to be consistent and comparable in traffic complexity between each other. This could be done by implementing a traffic complexity measure within the simulation environment, and keeping it consistent among the scenarios. There are several complexity metrics that have been developed, ranging from simple metrics such as aircraft density to computationally intensive ones. For the scope of the experiment at hand, in which a simplified ATC environment is simulated, the factors that were deemed most important are:

- **Aircraft density:** a factor directly related to workload, as a more dense sector requires a higher effort for supervising aircraft;
- **Aircraft velocity and heading variance:** a high variation in aircraft velocity or heading in the sector could lead to high supervisory workload;
- **Conflict density:** number of conflicts in a set time interval;
- **Conflict difficulty:** could be quantified by using the angle between two aircraft in conflict.

There are several metrics that include all these factors, but one that matches the level and scope of the experiment is the dynamic density metric developed by Wyndemere [64] for the National Aeronautics and Space Administration (NASA), which includes the influence of the above mentioned factors. Usually, other factors are included in complexity metrics as well, such as weather, aircraft characteristics or the presence of airports. The simplified nature of the simulation environment means that these are not applicable.

On the other hand, one subject will only work with two scenarios: the baseline case and one of the fictional aircraft cases. Therefore, memory retention will be low, especially as the scenarios span approximately one hour. Subjects will also be given a break in between scenarios to reset their level of vigilance, or even perform the runs on different days, depending on logistical flexibility. This opens up the possibility of using the same real traffic scenario for all cases, with modifications such as flipping and rotating the point of view over traffic, and tweaking the anomalies such that they do not appear similar. The experiment should also be conducted such that subjects do not have access to a clock or a timer to mitigate any pattern recognition in function of time. This way, the real traffic complexity measures will be the same for all scenarios, with required workload being influenced by the presence of fictional aircraft only.

A combination of these two options will be implemented within the experiment, as the complexity metric can be correlated with required workload, and thus with loss of vigilance and situation awareness in cases where workload is low. The methodology will be refined at a later stage of the project, when more information will be known about the state of the SARS-CoV-2 pandemic.

4.3.6. Variables and Measurement Methods

During the experiment, data will be collected through two methods: objective and subjective. Objective information consists of the digitally recorded parameters and metrics during each run of the experiment (such as subject score) whereas subjective data consists of questions answered by the subjects through surveys. The variables considered in the experiment are listed below according to their classification.

Independent Variables

- **Number of fictional aircraft:** will be defined as a percentage of the average number of real aircraft within the sector for a single run. This will be done in the future stages of the research project, as preliminary testing is required. The three levels for this variable are:
    - None (baseline case);
    - Low;
    - High.

Dependent Variables

- **Cognitive performance factors**, such as required workload, vigilance, situation awareness levels. Indicators for these will be obtained in the following way:
    - Workload: traffic complexity metric;

  - Vigilance: anomaly detection reaction time, manual control performance;
  - Situation awareness: anomaly detection rate, reaction time, manual control performance, fictional aircraft control performance, subjective survey.

- **Control performance parameters**, recorded throughout the simulation for both fictional and real traffic. These are:
  - Flown track miles;
  - On-track score;
  - Number of losses of separation;
  - Number of issued commands;
  - Number of mouse clicks;
  - Minimum distance between aircraft during a run;
  - Average distance between aircraft.

Controlled Variables

- **Aircraft type and characteristics:** all aircraft will be of the same type, and will therefore have the same characteristics.
- **Traffic complexity metric:** within all scenarios, real traffic will be characterised by the same complexity metric. The number of real/fictional aircraft will vary within a single run, but should be the same between runs.
- **Aircraft velocity:** aircraft will not experience change in velocities, and operators cannot give velocity commands. This is done to decrease the overall degrees of freedom, which can improve training time.

The subjective rating survey will be conducted after each run with the scope of obtaining a measure of the overall situation awareness level of the subject throughout the run. Due to their low level of intrusiveness and low resource requirement, the SASHA [65] and SART [66] questionnaires proposed by EUROCONTROL. For the situation at hand, due to more widespread use and documented validity of results, the SART survey will be used after each case. The survey consists of questions that seek to obtain an indication of the level of three domains during the run: attention demand, attention supply, and understanding experienced by the operator. The method is described and evaluated by Selcon and Taylor [67].

4.3.7. Subject Training and Briefing

Subject training is important in mitigating the effects of learning during the experiment, which could potentially result in an erroneous improvement in performance observed within the data. Thus subjects needs to be proficient enough with using SectorX when they enter the experiment that the learning effect is minimised. This can be attained through a thorough training process beforehand. There are several methods to achieve this, the most straightforward being to train participants just before their experimental runs. The subject would perform several training runs, in which they get used to manually controlling aircraft within the simulator, and how to control fictional aircraft when automation is functioning. However, this would probably increase the experiment time per subject significantly in combination with the lengthy runs, which in undesirable.

Another option would be off-site training, where participants receive a training program together with a briefing manual that they can use to train independently. The participants would then need to take a short test before the experiment runs to check if their level of training is adequate. The issue with this method would be that it is difficult to control the level of training that participants will achieve, as some might train more than others. This could mean that, for the between subjects part of the experiment (comparing the two fictional aircraft cases), it could produce statistical anomalies if the number of participants is low. A potential solution to this could be that the test is replaced by a required average score level. The participants would be instructed to train until they achieved a set average score over a defined time interval. Thus, all participants would receive similar levels of training.

The participants will be briefed both before training and before the experiment on how the simulator works and what tasks they need to perform during a run. Due to the ethical considerations, participants will be informed that anomalies and automation failure will occur during the experiment, as well as given a description and examples of what anomalies they can expect. Participants will not be given information about the time frame of these events, nor about the temporal structure of the scenarios.

4.4. Ethical Implications and Participant Motivation

As described previously in Chapter 2, gamification is prone to ethical implications, such as operator manipulation and lack of transparency. The experiment design seeks to mitigate these by informing the participants of the fictional nature of some aircraft on the screen, and the reasons why they are present. Furthermore, the participants will be aware that anomalies and failure will occur, thus avoiding inducing a high amount of stress. As subjects will not receive information about the structure of the experiment (number of cases they will perform, length of these cases), they will be informed of an estimated total time that will be required.

Participants will be motivated throughout a run through the use of the previously described scoring system. This enables the use of a reward based system, in which participants with high scores could receive a symbolic prize. It should be noted that the baseline case does not include a scoring system, and thus is not counted towards the reward. Thus, for the fictional aircraft cases, the high score threshold will have to be set differently for each of the two cases to account for the different number of fictional aircraft.

5

Conclusion

The aviation industry is increasingly turning towards automation as safer and more efficient solutions are developed. The air traffic control industry, known for its resilience to change, is adapting to the increased amount of traffic around the world through the implementation of tools that seek to aid controllers. Furthermore, automation has already been implemented within information acquisition systems in ATC, with further development in other areas expected in the future. The SESAR project estimates, in the European ATC Masterplan of 2020 [2], that a high level of automation will be attained by the year 2040. However, this implies that controllers will become system supervisors and thus will experience the negative effects of this role: loss of vigilance, low situation awareness and boredom.

The literature study revealed that a potential solution to this problem could be found in the use of gamification (game elements and concepts used within a non-game related situation). These kind of tools are currently being used in other domains to motivate humans performing highly monotonous tasks, such as luggage screening in airports. Here, threat image projection is used: fictional threats (knives, firearms) are places within the workflow, and operators need to identify them. This increases the number of threats that the operator experiences, thus increasing vigilance and performance.

The same concept is proposed for implementation within the air traffic control environment: fictional aircraft are introduced on the radar display of an air traffic controller that supervises an automatic conflict detection and resolution system. The fictional aircraft are differentiated from the real traffic visually (colour and shape), and they need to be guided through the sector to their exit waypoint by avoiding other fictional and real traffic. The aim of this process is to keep the operator engaged using a secondary task, while maintaining their attention within the work space. Furthermore, it is hypothesised that, because the operator engagement is maintained, manual control performance in case of automation failure will be better due to the less sudden transition from supervisory to manual control.

An experiment proposal has been developed in order to test the hypotheses. The experiment will test a simplified version of the aid tool using the SectorX ATC simulation environment developed by the Control & Simulation department of the Faculty of Aerospace Engineering of TU Delft. Students with prior knowledge of the air traffic control domain will be used as experiment subjects. The goal of the experiment is to determine if gamification can achieve better vigilance and situation awareness, and observe the cognitive effect on the operators. Due to the highly controlled experimental environment, the results of the experiment are expected to be reliable in determining whether to reject or accept the hypotheses.

Bibliography

[1] S. Becken and F. Carmignani. “Are the current expectations for growing air travel demand realistic?” In: *Annals of Tourism Research* 80 (2020), p. 102840. ISSN: 0160-7383. DOI: 10.1016/j.annals.2019.102840.

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

[3] U. Metzger and R. Parasuraman. “Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload”. In: *Human Factors* 47.1 (2005), pp. 35–49. ISSN: 0018-7208. DOI: 10.1518/0018720053653802.

[4] K. Volz et al. “An evaluation of cognitive skill degradation in information automation”. In: Proceedings of the Human Factors and Ergonomics Society. 2016, pp. 191–195. DOI: 10.1177/1541931213601043.

[5] S. Dixon and C.D. Wickens. “Automation Reliability in Unmanned Aerial Vehicle Control: A Reliance-Compliance Model of Automation Dependence in High Workload”. In: *Human factors* 48 (Feb. 1, 2006), pp. 474–86. DOI: 10.1518/001872006778606822.

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

[7] E. Sunil et al. “The Influence of Traffic Structure on Airspace Capacity”. In: *ICRAT 2016 - 7th International Conference on Research in Air Transportation*. Best paper award for the Network Management track. Philadelphia, United States, June 2016.

[8] SESAR Joint Undertaking. *A proposal for the future architecture of the European airspace*. MG-03-19-090-EN-C. doi:10.2829/718392. Luxembourg: Publications Office of the European Union, 2019. ISBN: 978-92-9216-116-3.

[9] Thomas B Sheridan and William L Verplank. *Human and computer control of undersea teleoperators*. Tech. rep. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, 1978.

[10] On-Road Automated Driving (ORAD) committee. *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. SAE International. DOI: 10.4271/J3016_201401. URL: https://www.sae.org/content/j3016_201401 (visited on 01/27/2020).

[11] A.R. Bestugin et al. *Air Traffic Control Automated Systems*. Springer Aerospace Technology. Singapore: Springer Singapore, 2020. ISBN: 978-981-13-9385-3. DOI: 10.1007/978-981-13-9386-0.

[12] T. Noskievič and J. Kraus. “Air Traffic Control Tools Assessment”. In: *MAD - Magazine of Aviation Development* 5.2 (Apr. 15, 2017), pp. 6–10. ISSN: 1805-7578. DOI: 10.14311/MAD.2017.02.01.

[13] *News ATS Ups Capacity in Karlsruhe – Airports International*. 2012. URL: https://airportsinternational.keypublishing.com/2012/02/06/news-ats-ups-capacity-in-karlsruhe/ (visited on 01/28/2020).

[14] S. Nieto and F. Javier. “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. ISSN: 0954-4100. DOI: 10.1177/0954410015596763.

[15] C. Nava Gaxiola et al. “Assessment of the North European free route airspace deployment”. In: *Journal of Air Transport Management* 73 (Oct. 1, 2018), pp. 113–119. ISSN: 0969-6997. DOI: 10.1016/j.jairtraman.2018.08.008.

[16] P. Tomasello et al. *STRESS: Final Project Results Report*. 00.03.00. Aug. 2018, p. 81.

[17] B. Berberian et al. “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.

[18] O. Ohneiser et al. “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.

[19] M. Bekier, B.R.C. Molesworth, and A. Williamson. “Tipping point: The narrow path between automation acceptance and rejection in air traffic management”. In: *Safety Science* 50.2 (2012), pp. 259–265. ISSN: 09257535. DOI: 10.1016/j.ssci.2011.08.059.

[20] R. Parasuraman and V. Riley. “Humans and Automation: Use, Misuse, Disuse, Abuse”. In: *Human Factors* 39.2 (June 1, 1997), pp. 230–253. ISSN: 0018-7208. DOI: 10.1518/001872097778543886.

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

[22] R.I. Thackray. *A comparison of detection efficiency on an air traffic control monitoring task with and without computer aiding.* American Psychological Association, 1989. DOI: 10.1037/e415342004-001.

[23] C.K. Foroughi et al. “Detecting automation failures in a simulated supervisory control environment”. In: *Ergonomics* 62.9 (2019), pp. 1150–1161. ISSN: 0014-0139. DOI: 10.1080/00140139.2019.1629639.

[24] R. Parasuraman, T.B. Sheridan, and C.D.D. Wickens. “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), pp. 286–297. ISSN: 10834427. DOI: 10.1109/3468.844354.

[25] M. Endsley. “Design and Evaluation for Situation Awareness Enhancement”. In: *Proceedings of the Human Factors Society Annual Meeting* 32.2 (Oct. 1, 1988), pp. 97–101. ISSN: 0163-5182. DOI: 10.1177/154193128803200221.

[26] M. Endsley. “Toward a Theory of Situation Awareness in Dynamic Systems”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (Mar. 1995), pp. 32–64. ISSN: 0018-7208, 1547-8181. DOI: 10.1518/001872095779049543.

[27] E. Rovira and R. Parasuraman. “Transitioning to Future Air Traffic Management: Effects of Imperfect Automation on Controller Attention and Performance”. In: *Human factors* 52 (June 1, 2010), pp. 411–25. DOI: 10.1177/0018720810375692.

[28] R. Parasuraman, T.B. Sheridan, and C.D. Wickens. “Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs”. In: *Journal of Cognitive Engineering and Decision Making* (June 1, 2008). DOI: 10.1518/155534308X284417.

[29] J. Mercer et al. “Impact of Automation Support on the Conflict Resolution Task in a Human-in-the-Loop Air Traffic Control Simulation”. In: *IFAC-PapersOnLine*. 13th IFAC Symposium on Analysis, Design, and Evaluation ofHuman-Machine Systems HMS 2016 49.19 (Jan. 1, 2016), pp. 36–41. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2016.10.458.

[30] A. Nunes. “The Impact of Automation Use on the Mental Model: Findings from the Air Traffic Control Domain”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47.1 (Oct. 1, 2003), pp. 66–70. ISSN: 2169-5067. DOI: 10.1177/154193120304700114.

[31] M. Endsley and M.D. Rodgers. “Distribution of Attention, Situation Awareness and Workload in a Passive Air Traffic Control Task: Implications for Operational Errors and Automation”. In: *Air Traffic Control Quarterly* 6.1 (Jan. 1, 1998), pp. 21–44. ISSN: 1064-3818. DOI: 10.2514/atcq.6.1.21.

[32] J.J. Cañas et al. “Mental Workload in the Explanation of Automation Effects on ATC Performance”. In: *Human Mental Workload: Models and Applications*. Ed. by L. Longo and M. C. Leva. Communications in Computer and Information Science. Cham: Springer International Publishing, 2019, pp. 202–221. ISBN: 978-3-030-14273-5. DOI: 10.1007/978-3-030-14273-5_12.

[33] J. S. Warm, R. Parasuraman, and G. Matthews. “Vigilance Requires Hard Mental Work and Is Stressful”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50.3 (June 2008), pp. 433–441. ISSN: 0018-7208, 1547-8181. DOI: 10.1518/001872008X312152.

[34] L. Buck. “Reaction time as a measure of perceptual vigilance.” In: *Psychological Bulletin* 65.5 (1966), pp. 291–304. ISSN: 1939-1455, 0033-2909. DOI: 10.1037/h0023207.

[35] V. Pop et al. “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.

[36] R. Parasuraman, R. Molloy, and I.L. Singh. “Performance Consequences of Automation-Induced ’Complacency’”. In: *The International Journal of Aviation Psychology* 3.1 (Jan. 1, 1993), pp. 1–23. ISSN: 1050-8414. DOI: 10.1207/s15327108ijap0301_1.

[37] C.D. D. Wickens et al. “Complacency and Automation Bias in the Use of Imperfect Automation”. In: *Human Factors* 57.5 (Aug. 1, 2015), pp. 728–739. ISSN: 0018-7208. DOI: 10.1177/0018720815581940.

[38] J. Lee and N. Moray. “Trust, control strategies and allocation of function in human-machine systems”. In: *Ergonomics* 35 (May 31, 2007), pp. 1243–1270. DOI: 10.1080/00140139208967392.

[39] Christopher Wickens. “Attentional Tunneling and Task Management”. In: *2005 International Symposium on Aviation Psychology* (Jan. 1, 2005), pp. 812–817. URL: https://corescholar.libraries.wright.edu/isap_2005/121.

[40] S. T. Shorrock. “Errors of perception in air traffic control”. In: *Safety Science* 45.8 (Oct. 1, 2007), pp. 890–904. ISSN: 0925-7535. DOI: 10.1016/j.ssci.2006.08.018.

[41] A. Haslbeck and H. Hoermann. “Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills Among Airline Pilots”. In: *Human Factors* 58.4 (June 1, 2016), pp. 533–545. ISSN: 0018-7208. DOI: 10.1177/0018720816640394.

[42] B. Kirwan. “The role of the controller in the accelerating industry of air traffic management”. In: *Safety Science* 37.2 (Mar. 1, 2001), pp. 151–185. ISSN: 0925-7535. DOI: 10.1016/S0925-7535(00)00047-3.

[43] M. Bekier, B.C. Molesworth, and A. Williamson. “Why Air Traffic Controllers Accept or Refuse Automated Technology”. In: *16th International Symposium on Aviation Psychology* (2011), pp. 615–620.

[44] C. Westin, C. Borst, and B. Hilburn. “Strategic Conformance: Overcoming Acceptance Issues of Decision Aiding Automation?” In: *IEEE Transactions on Human-Machine Systems* 46.1 (Feb. 2016), pp. 41–52. ISSN: 2168-2305. DOI: 10.1109/THMS.2015.2482480.

[45] R. Meuter and P. Lacherez. “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. ISSN: 0018-7208. DOI: 10.1177/0018720815616306.

[46] ESD. *HYDEN XP X-Ray Training overview – ESD*. URL: http://esdevelopment.co.za/?page_id=39 (visited on 02/13/2020).

[47] F. Hofer and A. Schwaninger. “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.

[48] A. Schwaninger et al. “The Impact of Image Based Factors and Training on Threat Detection Performance in X-ray Screening”. In: *International Conference on Research in Air Transportation*. 2008, p. 8. DOI: 10.13140/RG.2.1.1299.3526.

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

[50] E. Paja et al. *PACAS: A Gamified Platform for Participatory Change Management in ATM Systems*. Joint Proceedings of REFSQ-2018 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track co-located with the 23rd International Conference on Requirements Engineering: Foundation for Software Quality (REFSQ 2018), Utrecht, The Netherlands, March 19, 2018. 2018.

[51] F. Khatib et al. “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. ISSN: 1545-9993. DOI: 10.1038/nsmb.2119.

[52] M. Vansteenkiste and E. Deci. “Competitively Contingent Rewards and Intrinsic Motivation: Can Losers Remain Motivated?” In: *Motivation and Emotion* 27.4 (2003), p. 27.

[53] T. Kim and K. Werbach. “More than just a game: ethical issues in gamification”. In: *Ethics and Information Technology* 18.2 (June 2016), pp. 157–173. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/s10676-016-9401-5.

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

[55] C.D. Wickens, S. Gordon, and Y. Liu. *An introduction to human factors engineering*. New York: Longman, 1998. ISBN: 0-321-01229-1.

[56] K.J. Vicente and J. Rasmussen. “A Theoretical Framework for Ecological Interface Design”. In: *RISØ-M-2736* (1988).

[57] J. Rasmussen. “Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models”. In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13.3 (May 1983), pp. 257–266. ISSN: 0018-9472, 2168-2909. DOI: 10.1109/TSMC.1983.6313160.

[58] C. Borst et al. “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. ISSN: 2168-2305. DOI: 10.1109/THMS.2019.2919742.

[59] INDRA. *Air Traffic Control Automation System*. URL: https://www.indracompany.com/sites/default/files/atc_automationsyst_en_0.pdf (visited on 04/23/2020).

[60] U. Ahlstrom and L. Arend. “Color Usability on Air Traffic Control Displays”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 49.1 (Sept. 1, 2005). Publisher: SAGE Publications Inc, pp. 93–97. ISSN: 2169-5067. DOI: 10.1177/154193120504900121.

[61] N. Mackworth. “The Breakdown of Vigilance during Prolonged Visual Search”. In: *Quarterly Journal of Experimental Psychology* (1948). DOI: 10.1080/17470214808416738.

[62] P. Hancock. “In Search of Vigilance The Problem of Iatrogenically Created Psychological Phenomena”. In: *The American psychologist* 68 (Oct. 22, 2012). DOI: 10.1037/a0030214.

[63] M. Kim. “Enhancement of Air Traffic Controller’s Task Engagement for Smooth Transition from Supervisory to Manual Control”. MSc Thesis. 2019. URL: http://resolver.tudelft.nl/uuid:7fc69484-2c56-4811-9f20-5357bc8677c5 (visited on 04/24/2020).

[64] Wyndemere. *An Evaluation of Air Traffic Control Complexity*. Final Report NAS2-14284. 1996.

[65] EUROCONTROL. *Situation Awareness for SHAPE (SASHA) | HP repository*. 2012. URL: https://ext.eurocontrol.int/ehp/?q=node/1609 (visited on 04/24/2020).

[66] EUROCONTROL. *Situation Awareness Rating Technique (SART) | HP repository*. 2012. URL: https://ext.eurocontrol.int/ehp/?q=node/1608 (visited on 04/24/2020).

[67] S.J. Selcon and R.M. Taylor. “Evaluation of the Situational Awareness Rating Technique(SART) as a tool for aircrew systems design”. In: *AGARD, Situational Awareness in Aerospace Operations 8 p(SEE N 90-28972 23-53)* (1990).

A

Interface Design Concept (Graded)

The research project presented in this report focuses on the cognitive effects of the implementation of gamification within an ATC environment. Thus, there is less emphasis on the design of the user interface of the proposed tool. However, a preliminary design concept of the display elements and visualisation was created and presented below. It should be noted that the colours presented below are not accurate, the actual tool will be implemented within an ATC environment, which usually uses a dark background.

First of all, Figure A.1 portrays the nominal situation in which automation functions as expected and operator fulfils a supervisory role. There is an indicator that shows that automation is active, displaying information through colour and text. The air traffic controller has the option to request manual control at any time by using the provided button. Fictional aircraft are present on the screen. Note that in these examples, only the fictional aircraft tool is displayed, other ATC tools are not included.

The second portrayed situation is the case in which the operator requested manual control by pressing the button, shown in Figure A.2. This could be as a result of undetected automation failure or anomaly. In this case, the indicator shows that the automation is not functioning, and fictional aircraft disappear from the screen to allow focus on real traffic.

Finally, Figure A.3 portrays the situation in which the system detected the failure of automation and requests the intervention of the operator. Automation is therefore turned off, and an alert is displayed on the screen, in combination with audio signals. This is done through the use of bright colours for high visibility. The alert disappears once the first commands have been issued to aircraft, or once it is dismissed by the operator.

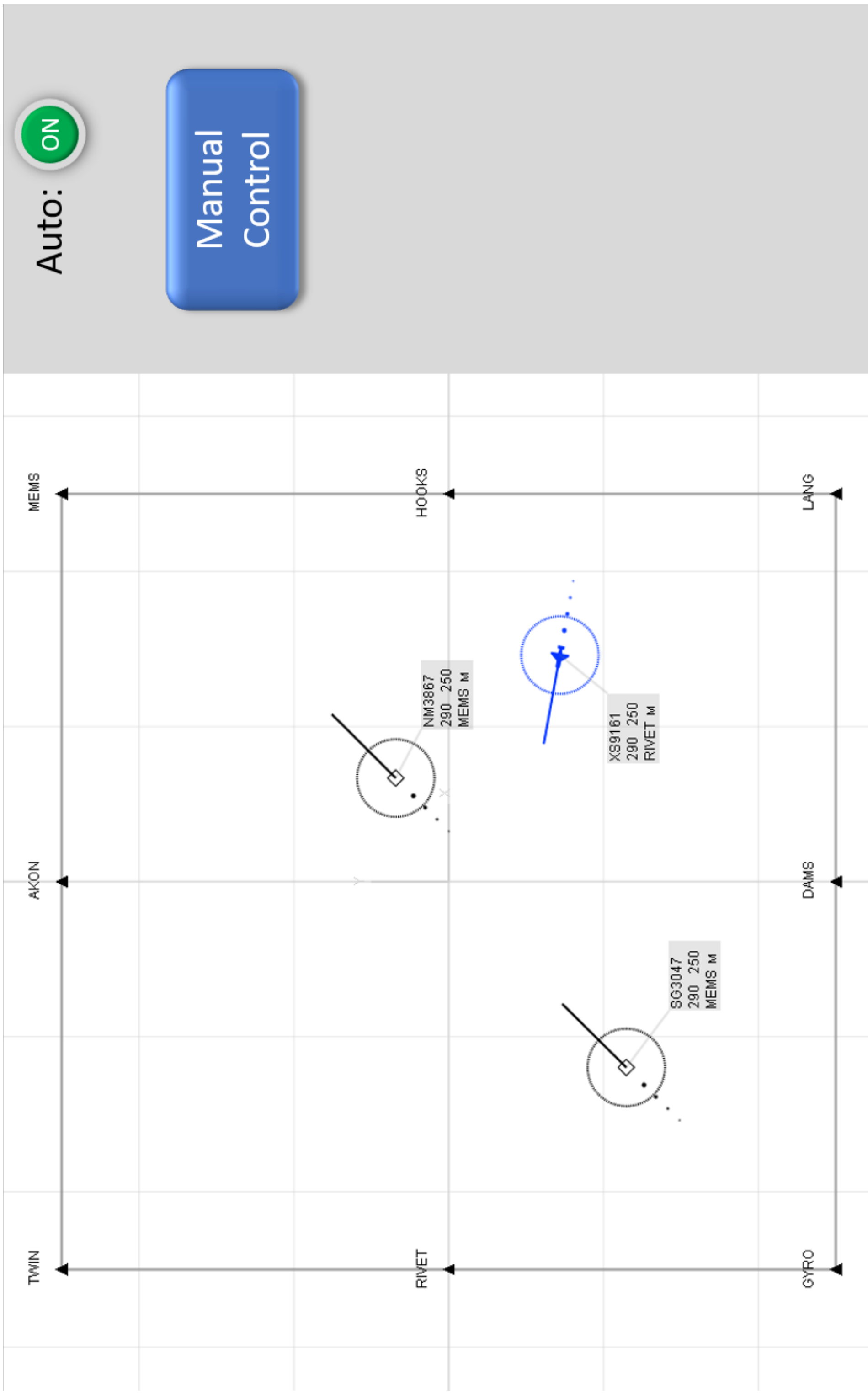


Figure A.1: User interface while automation is active.

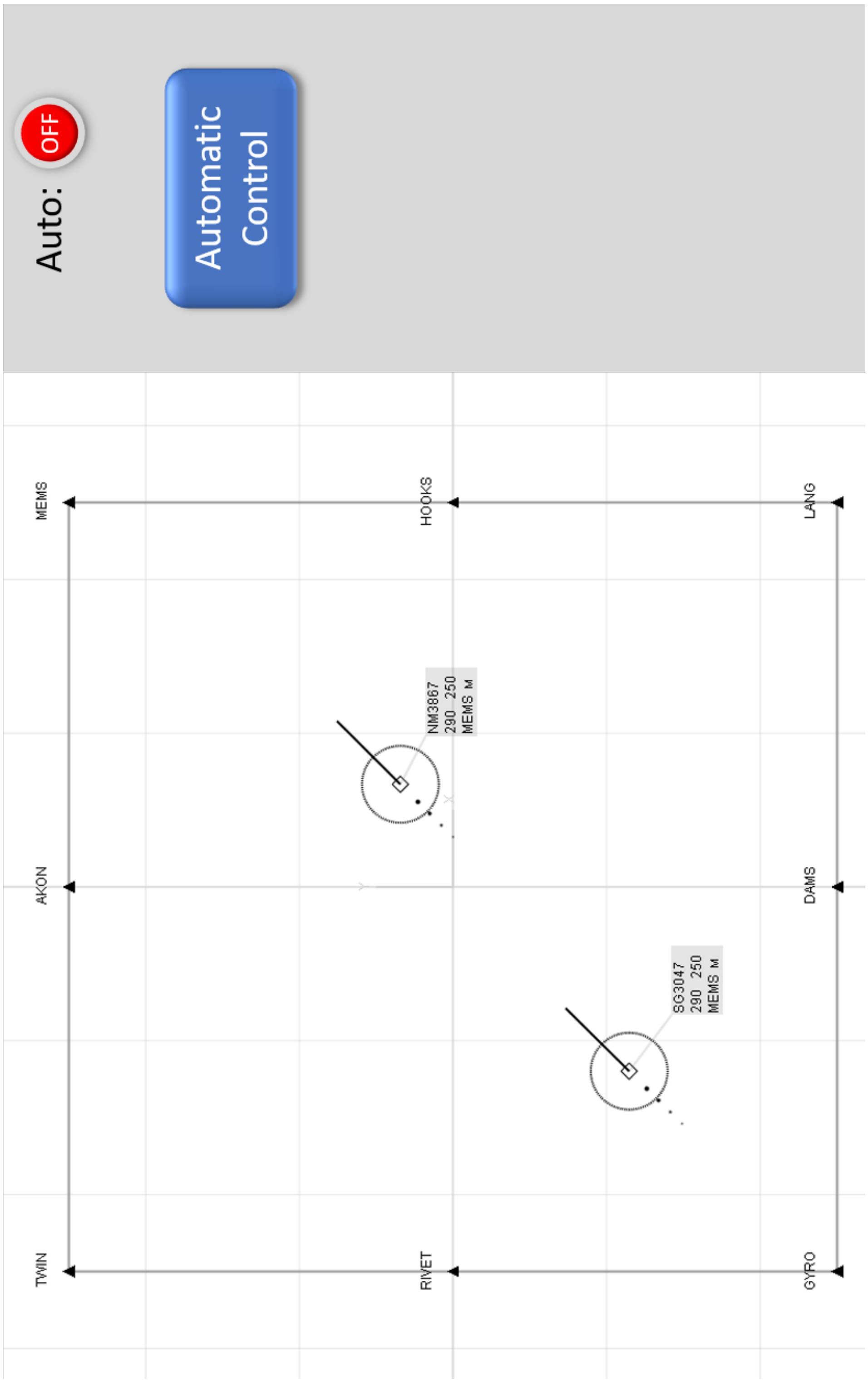


Figure A.2: User interface after manual control has been requested by operator. The button changes function, and can be used to re-engage automation.

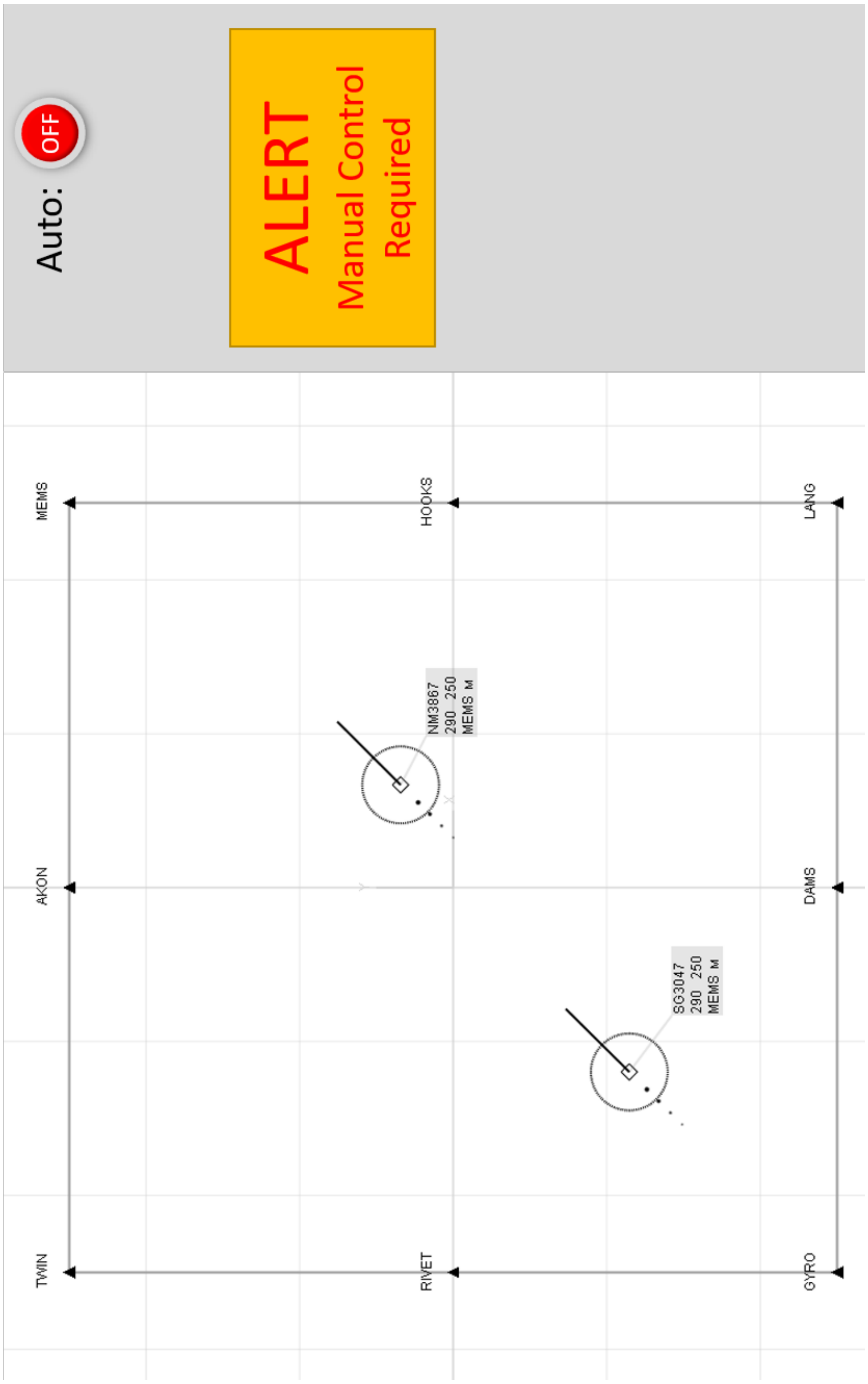


Figure A.3: User interface in case automation failure is detected by the system and manual control intervention is requested. The manual control button is not present in this state. After the alert has been dismissed, the ”automatic control” button, as presented in Figure A.2 is shown again.

III

Additional Appendices

Not graded

B

Experiment Procedure

Due to the conditions brought by the coronavirus pandemic, the experiment had to be conducted in a controlled environment, and most of the participant briefing was done through the manual presented in Appendix C. This document was sent between two and five days in advance by email to each participant. All participants confirmed reading the manual before the experiment.

All equipment was disinfected using ethanol solution before and after each experiment session. The participants were first asked to fill in a consent form, presented in Appendix D. Afterwards, all participants undertook the same training scenarios (TS), presented below:

TS 1 An introductory scenario that sought to act as a tutorial for the basic functionality of SectorX. It contained two aircraft that needed to be routed to their exit waypoints, and no conflicts. The initial position of aircraft is shown in Figure B.1.

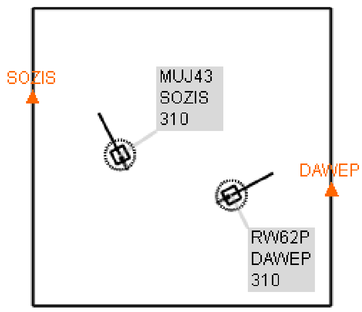


Figure B.1: Initial position of aircraft in Training Scenario 1.

TS 2 A scenario that introduces participants to conflicts and resolution manoeuvres, presented in Figure B.2.

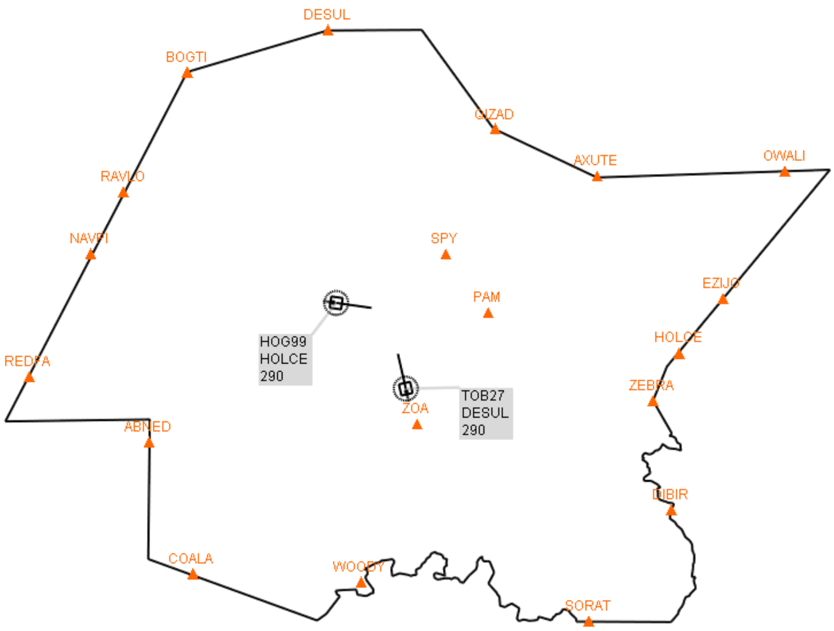


Figure B.2: Initial position of aircraft in Training Scenario 2.

TS 3 A simple scenario that contains two aircraft conflicts. The second conflict involves aircraft that have different indicated air speeds. The starting position of the scenario is presented in Figure B.3.

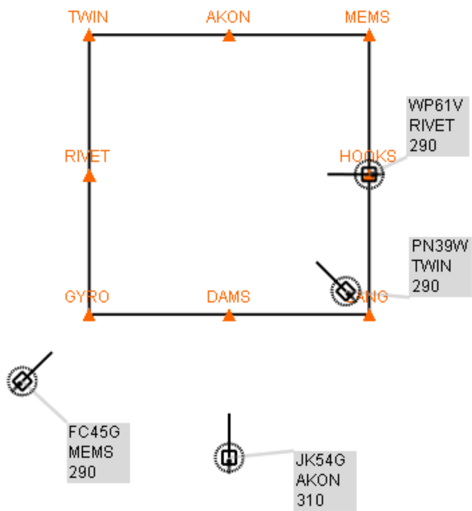


Figure B.3: Initial position of aircraft in Training Scenario 3.

TS 4 A more complex scenario in which main traffic flows are introduced, as shown in Figure B.4. Participants need to both de-conflict aircraft and route them to their correct exit waypoint.

TS 5 Another scenario of higher difficulty with a different traffic pattern, presented in Figure B.5.

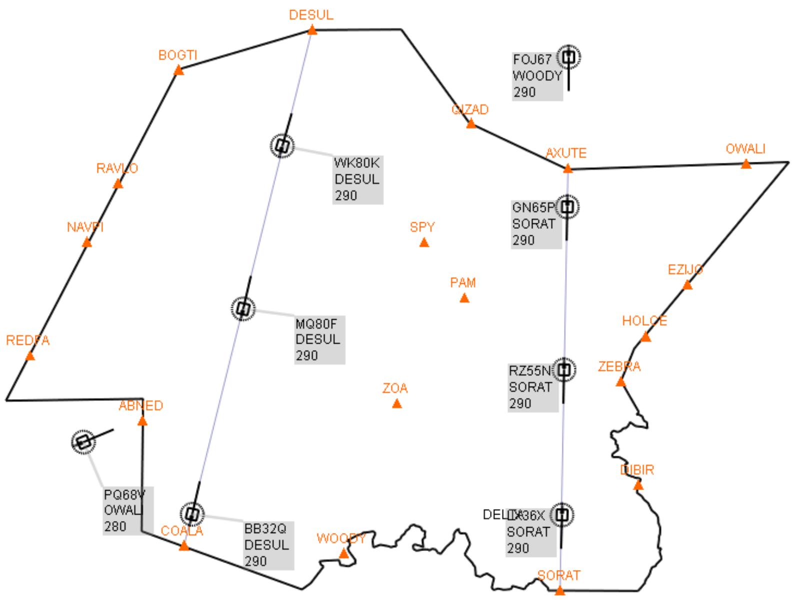


Figure B.4: Initial position of aircraft in Training Scenario 4.

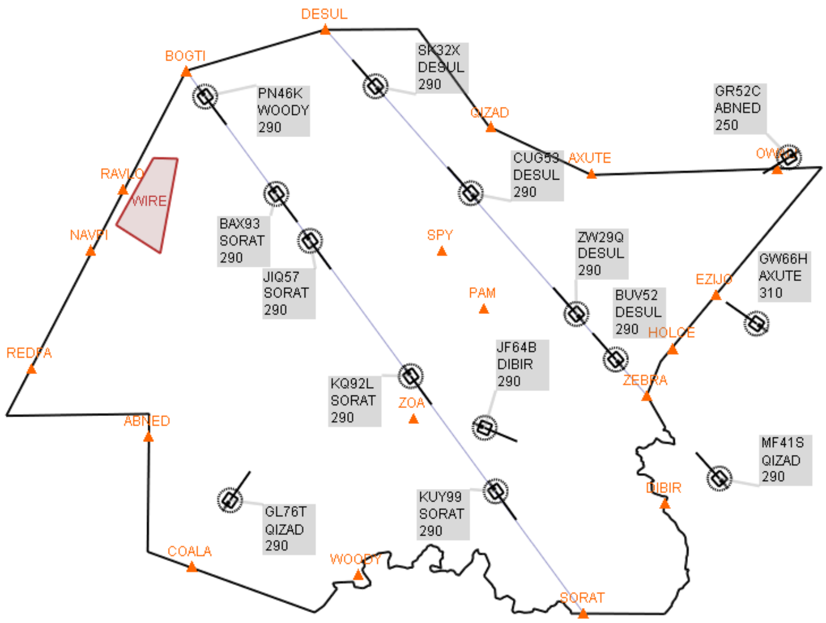


Figure B.5: Initial position of aircraft in Training Scenario 5.

TS 6 This scenario contains the same aircraft as Training Scenario 3, but real aircraft are controlled by automation, and fictional aircraft are present and must be manually controlled, as presented in Figure B.6.

TS 7 This scenario contains the same traffic as Training Scenario 5, with automation enabled for real aircraft and fictional aircraft present in the sector, as presented in Figure B.7.

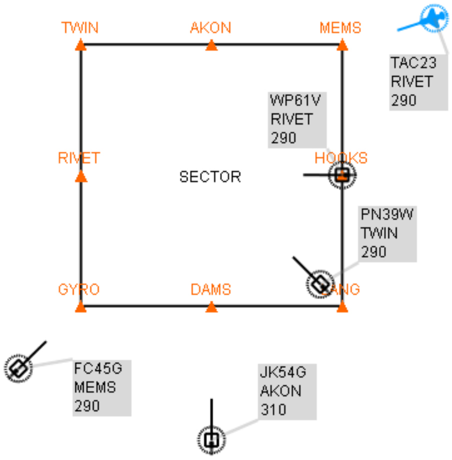


Figure B.6: Initial position of aircraft in Training Scenario 6.

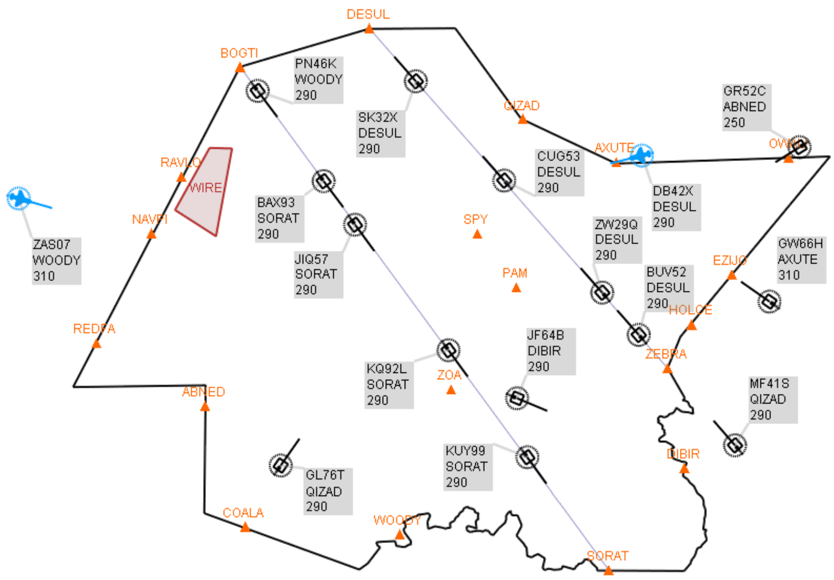


Figure B.7: Initial position of aircraft in Training Scenario 7.

The training procedure lasted for approximately 70 minutes, after which there was a break of 10 minutes. Then, the participant was informed which group they were assigned to. Participants in the baseline group received the following briefing:

> You are an air traffic controller that is tasked with supervising an automated conflict detection and resolution program. You must ensure aircraft are safely and efficiently routed towards their exit waypoints. Anomalies are known to occur, thus you must report and describe all automation anomalies you observed. At some point during the scenario, automation will fail, and manual control will be required. A failure notification will appear on screen, and you must dismiss it as soon as possible and take over manual control. You must then route aircraft safely

> and efficiently towards their exit waypoints until the end of the scenario. Automation will not be re-enabled.

On top of the briefing given to baseline group participants, fictional aircraft group participants also received the following briefing:

> On top of the supervision task, you must also route fictional aircraft safely and efficiently through the sector as if they were real aircraft. Keep in mind that automation is not aware fictional aircraft are present on screen.

The participant was then asked to remove any distracting elements (watch, phones), and then the corresponding traffic scenario was started. During the scenario, the researcher did not communicate with the participant in order to minimise distractions, unless the participant asked questions. Finally, after the scenario finished, the participants were invited to fill in the post-experiment survey, presented in Appendix E.

C

Experiment Briefing Manual

Participant Briefing and Training

By Calin Andrei Badea

1. Introduction

Automation has proven to be a good tool for achieving better safety and efficiency in air travel. One domain where its implementation has been slow paced is air traffic control, however efforts are being made to achieve a high level of automation by the year 2040. This implies that air traffic will be controlled automatically, while air traffic controllers will be given a supervisory role: detecting exceptions that might occur. The issue with this is that humans are not particularly good at performing supervisory tasks when unaided. Even highly trained professionals experience fatigue and boredom, especially when the task is monotonous and lacks engagement.

One potential solution to this could be found within the concept of gamification: the use of game elements and techniques in non-gaming contexts with the purpose of increasing engagement and motivation. A spin-off of this technique, threat image projection (TIP) is already being used within the aviation industry to increase the performance of luggage screeners at airports. It consists in the occasional placement of fictional dangerous items, such as firearms, within the work flow of an operator, as shown in Figure 1. This increases the rate of occurrence of threatening objects, thus increasing operator vigilance and maintaining their alertness.

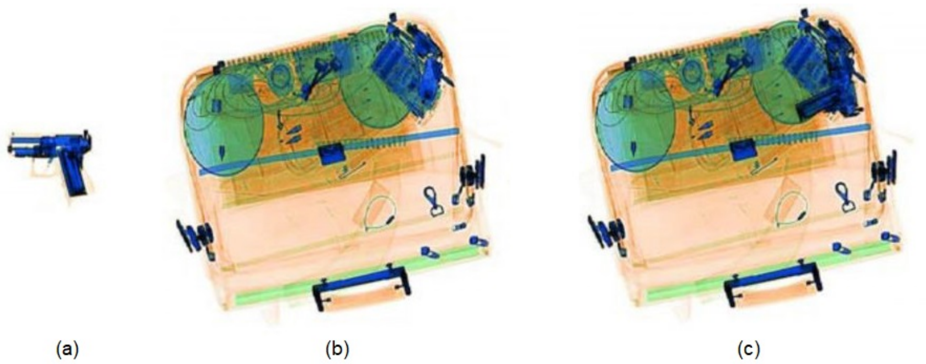


Figure 1: Threat image projection in luggage screening: random fictional threats (a) are imposed on top of random fictional luggage (b). Final product (c) is introduced within the work flow of the X-ray machine operator.

The scope of the experiment at hand is to research whether applying this concept within a highly automated air traffic control environment achieves the same improvements in operator cognitive performance for anomaly and failure detection. Fictional aircraft are displayed among actual traffic, the latter being fully controlled by automation. The operator (you) has to manually control the fictional aircraft as if there are real, guiding them safely and efficiently to their destination using the minimum amount of track miles (distance travelled).

2. SectorX ATC Simulator

The following section presents the simulation environment that will be used for the experiment.

2.1. General Functionality

SectorX is an ATC simulator developed within the Control and Simulation department. A modified version of it that implements the proposed fictional aircraft concept will be used during the experiment. An example of the simulation screen is shown in Figure 2. The airspace sector is delimited by a light blue line, and waypoints are represented as triangles, with their designated name written above them. The waypoints that sit close to the sector boundary are the exit/entry points.

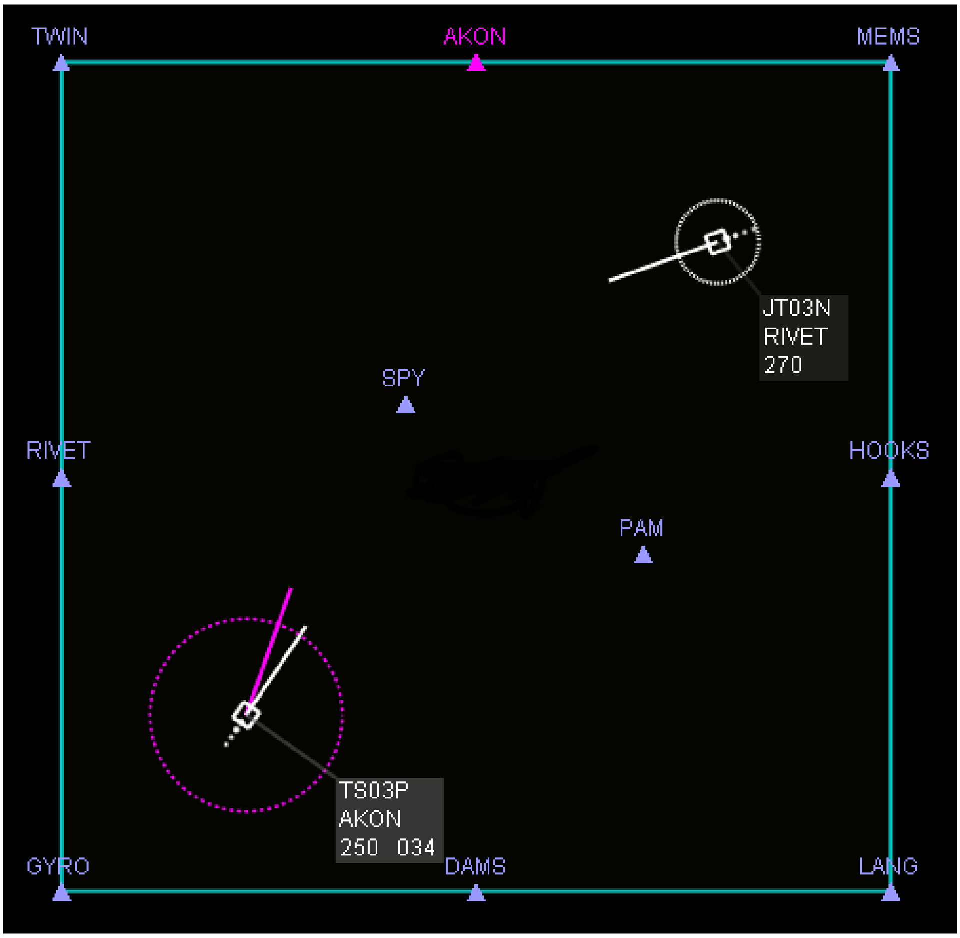


Figure 2: Example of SectorX simulator screen.

In this example, there are two aircraft present in the sector, represented by white, square icons. If an aircraft is selected by the user by clicking on it, it enters command mode (TS03P), and can be deselected by pressing anywhere else on the screen. Aircraft direction and velocity are represented by the heading and length of the white solid line. When not selected, aircraft have a circle around them with a radius of 2.5 nautical miles which helps enforce the 5 nautical mile separation minimum between two aircraft (in other words, the circles of two different aircraft must not intersect). The history dots behind each aircraft show the previous positions: 5 seconds, 10 seconds and 15 seconds prior to present, as the radar screen updates once every 5 seconds.

When an aircraft is selected, the user is shown the required heading for the exit waypoint in the form of a magenta line (which is also highlighted, see Figure 2, TS03P is heading to AKON). The aircraft labels also show different information depending on whether the aircraft is selected or not. If not selected, the aircraft label shows the **aircraft ID (TS03P)**, the **exit waypoint (AKON)**, and the **aircraft velocity in knots (250 kts)**. If the aircraft is selected, the label also shows the aircraft **heading in degrees (034 degrees)**.

Aircraft can be controlled by selecting them and issuing a heading command. During this experiment, the commands are limited to heading commands, which can be given by first selecting an aircraft, clicking within the command circle in the desired direction of travel, and then pressing "Enter" on the keyboard, as shown in Figure 3. The thicker magenta line represents the required heading to reach the exit waypoint AKON. Furthermore, the aircraft in the experiment will all **fly at the same flight level** (altitude), therefore the experiment occurs in a 2D environment.

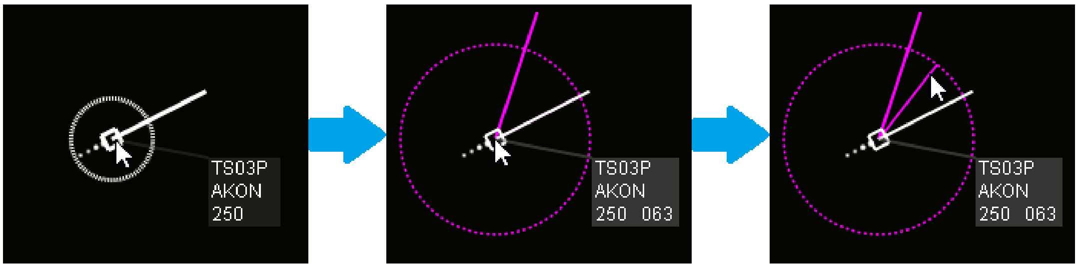


Figure 3: Procedure to give a heading command to an aircraft: move the mouse on top of the aircraft, select it by clicking once, then select the desired heading by clicking once within the pink circle. Finally, press enter to apply.

2.2. Fictional Aircraft

As previously mentioned, fictional aircraft were implemented within SectorX for the purpose of this experiment. These behave like actual aircraft, but are graphically different, with a different icon as well as different colour, as shown in Figure 4.

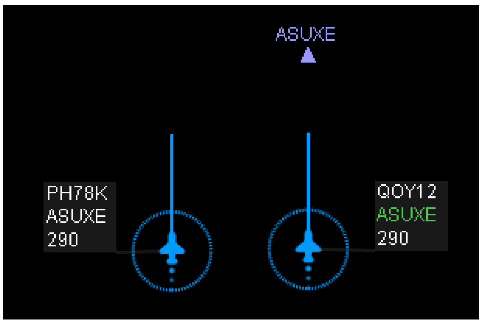


Figure 4: Fictional aircraft; aircraft on right is heading directly towards its designated waypoint, aircraft on left is not.

Fictional aircraft are represented by an aircraft-like icon and the colour blue. If a fictional aircraft is heading directly towards its exit waypoint, the waypoint name in the label turns green. **Note that this does not apply to real aircraft.**

2.3. Automation Features

SectorX is also capable of an automated control mode, in which traffic is routed and organised automatically. In this mode, the human operator can only control fictional aircraft (if present), and must supervise the actions of the automated algorithm. Automation does not control nor avoid fictional aircraft, and it is the responsibility of the operator to prevent conflicts between real and fictional aircraft. When a real aircraft receives a command and performs a manoeuvre, the colour of its label colour changes to indicate this, as shown in Figure 5.

Figure 5: Aircraft label colour changes when aircraft received a command and is manoeuvring.

Part of the supervision task is reporting any anomalies that might occur in the automated algorithm. This can be done by clicking the aircraft considered to be inappropriately handled by automation. If automation is currently running (indicated visually in the top left corner of the screen, see Figure 6), the report anomaly window will open, as shown in Figure 7. On this screen, a description of the anomaly can be written and sent. If the aircraft selection was performed accidentally, the screen can be dismissed without reporting.

Figure 6: Automation on indicator, is displayed in the top left corner of the screen.

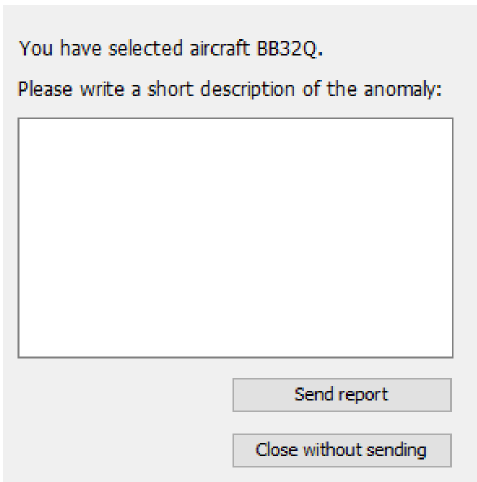


Figure 7: Anomaly report window, appears on the left side of the screen if a real aircraft is selected while automation is running.

On rare occasions, automation can also experience a catastrophic failure. In this case, the controller is required to intervene and take over control manually of all traffic. SectorX

Figure 8: Automation indicator when automation is off.

Figure 9: Alert box displayed when automation turns off.

will let the operator know through a flashing alert and the automation indicator showing that it turned off, as shown in Figures 8 and 9. When this happens, the controller needs to dismiss the alert as soon as possible **by clicking on it** to be able to start controlling aircraft manually. Fictional aircraft disappear from the screen as soon as the alert is issued.

3. Experimental Setup and Goals

The following section presents the format of the experiment, and what is expected from participants. The experiment will last 3 to 4 hours. The first part of the experiment is the training phase, where you will be instructed and trained in all aspects of using the SectorX simulator. This will last for approximately 1 hour and 15 minutes, after which there will be a 15 minute break.

The measurement part of the experiment will last for a variable amount of time for each participant. At first, automation will be controlling the actual traffic, and you will have to supervise its actions. In this time, you will have to control fictional aircraft (if present), and report any anomalies that you might encounter. These anomalies are not critical, and thus the algorithm will continue functioning even after one or several have occured. The possible anomalies are known to be the following, presented in Figure 10:

- **Loss of separation:** the 5 nautical mile separation minimum between two real aircraft is breached and the 2.5 nautical mile circles around them overlap;
- **Breach of restricted airspace:** a real aircraft breaches the minimum separation for restricted airspace, and is considered an anomaly **if the 2.5 nautical mile circle around a real aircraft overlaps with a restricted airspace sector** (indicated as a red polygon);
- **Wrong exit waypoint:** if a real aircraft is directed towards the wrong exit waypoint.

It is important to mention that anomalies might not be immediately evident, as the automated algorithm will sometimes perform actions to prevent future conflicts. If you report an anomaly but in the end it did not occur, you can always send a second report mentioning this. It is preferred that you send a report when you are convinced that what you witnessed was an anomaly. Furthermore, the automated algorithm will try to use the

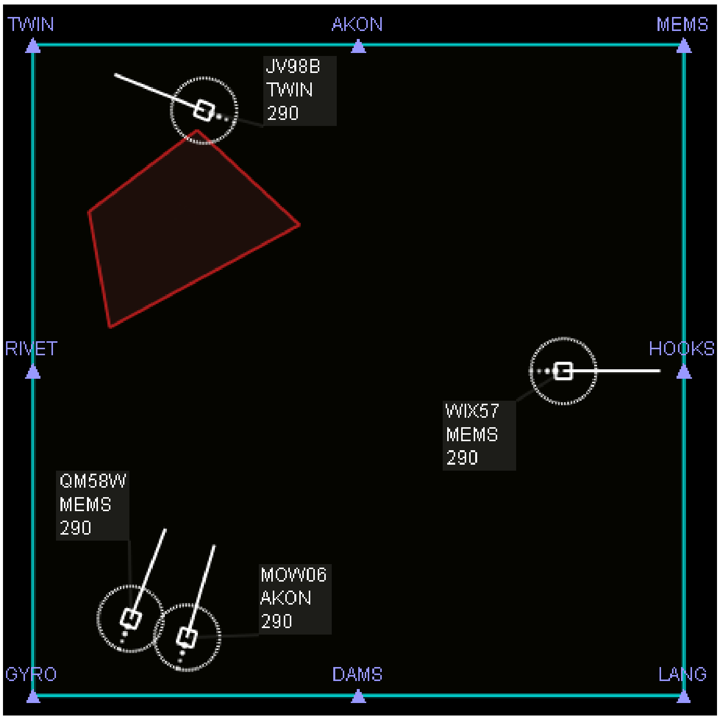


Figure 10: Anomalies known to occur when automation is active: QM58W and MOW06 experienced loss of separation, JV98B breached the 2.5 nautical mile separation minimum for restricted sectors, and WIX57 is heading towards the wrong exit waypoint.

main routes (shown as lines in Figure 11) in the sector as much as possible, but will divert aircraft from them to prevent future conflicts. This is considered normal behaviour.

At a random time during the experiment, automation will experience a catastrophic failure. When this occurs, you must acknowledge and dismiss the notification by clicking on the alert as soon as possible and take over manual control of the aircraft. Fictional aircraft will disappear from the screen. Your mission is to control real aircraft as efficiently and safely as possible, and direct them to their exit waypoint using the minimum amount of track miles (distance travelled).

The experiment participants will be divided into two groups: one that will have fictional aircraft on screen, and a control group that will not, and will have to supervise automation without the use of fictional aircraft. You will be informed during training which group you are part of.

The sector that will be used during the experiment is presented in Figure 11. It is a cropped version of the Delta sector controlled by the Maastricht Upper Area Control. Please familiarise yourself with the waypoint placement and names, the restricted area placement (red sector boundaries), the main air routes and shape of the sector.

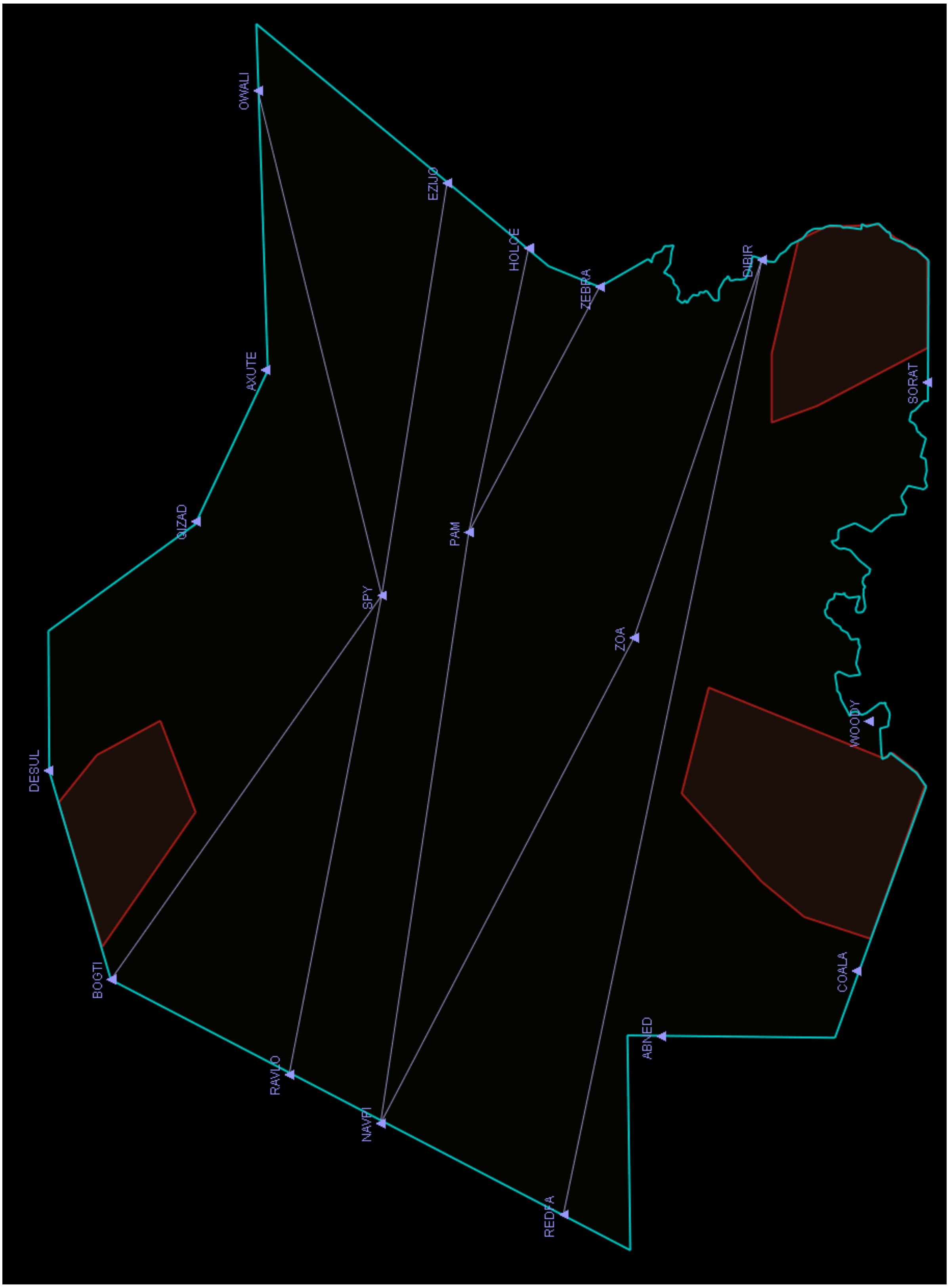


Figure 11: The sector that will be used during the experiment.

4. Best Practices and Tips

The following section provides you with some tips and tricks that you should use during the experiment.

- When a conflict involves two aircraft with different velocities, try routing the slower aircraft behind the faster one as it is generally the quickest way to solve a conflict.
- A good way to solve a conflict is to route one of the aircraft directly towards the current position of the other. This way, you will route one aircraft behind the other and solve the conflict relatively efficiently, as shown in Figure 12. This will work less well if the aircraft are already close to each other.

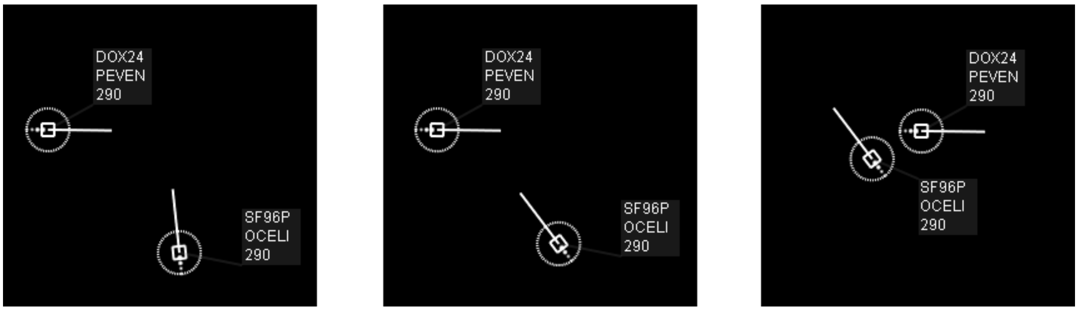


Figure 12: Simple way to solve a conflict between two aircraft: route one of the aircraft towards the present position of the other.

- In case of a possible conflict, it is better to act sooner and preventively than later. Aircraft turn relatively slowly, thus change their heading with difficulty, so if you wait until the last moment it might be more difficult to solve the conflict.
- Try to find solutions that minimise the number of commands you give in total, as this will keep your workload low (and that of the pilots as well), and might result in better efficieny overall.
- Safety always comes first, even if this implies an aircraft must take a less direct route towards its exit waypoint. This means two things: if automation does not route an aircraft directly to its exit waypoint when expected, it might be that a conflict is being avoided through this; furthermore, when manually controlling aircraft, do not hesitate to choose the safer solution over a less certain, more direct route.

D

Experiment Consent Form

Consent Form for “Gamification: Improving Supervisory Control Performance in ATC”

Please tick the appropriate boxes	Yes	No
Taking part in the study		
I have read and understood the experiment information and briefing dated [DD/MM/YYYY], or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction.	○	○
I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason.	○	○
I understand that taking part in the study involves having performance data automatically stored in an anonymous manner when completing the experiment.	○	○
I understand that taking part in the study involves me answering questions to surveys.	○	○
I understand that taking part in the study involves being subjected to stressful situations.	○	○
Use of the information in the study		
I understand that information I provide will be used for in the paper and thesis report on an anonymous basis.	○	○
I understand that personal information collected about me that can identify me, such as my name, email address, and phone number, will not be shared beyond the study team. *The study team does not aim to collect any personal information.*	○	○
I agree that my information can be quoted in research outputs on an anonymous basis. *The study team does not aim to collect any personal information.*	○	○
Future use and reuse of the information by others		
I give permission for the recorded performance data, answers to surveys, and video recordings, that I provide to be archived in secure folders so it can be used for future research and learning. All data is stored anonymous. Access is safeguarded and not to be used for commercial use. *The study team does not aim to collect any personal information.*	○	○
Health Risks		
I understand that the study team and the Delft University of Technology is not responsible for any mental and or physical damage incurred.	○	○
I understand that the study team and the Delft University of Technology is not responsible for any implications regarding COVID-19 despite the taken measures.	○	○
COVID-19		
I confirm that the researcher has provided me with detailed safety instructions to ensure my experiment session can be performed in line with current RIVM COVID-19 regulations at all times and that these instructions are fully clear to me.	○	○
I understand that also for my travel to/from the experiment session I should at all times adhere to the current RIVM COVID-19 regulations.	○	○

Signatures

______________________ ___________________ ________
Name of participant Signature Date

I have accurately read out the information sheet to the potential participant and, to the best of my ability, ensured that the participant understands to what they are freely consenting.

______________________ ___________________ ________
Researcher name Signature Date

Contact details:

Name: Andrei Badea

Phone Number: +40728873657

Email Address: c.a.badea@student.tudelft.nl

E

Post Experiment Survey

The survey participants filled in after the experiment run is presented in this chapter. Participants in the baseline group were not given questions targeted towards participants in the fictional aircraft group.

Post Experiment Survey

Please answer the questions below to the best of your ability.

What is your subject number?

How changeable was the situation? Was it highly unstable and likely to change suddenly or stable and straightforward?

	1	2	3	4	5	6	7	
Unstable	○	○	○	○	○	○	○	Stable

How complicated was the situation? Was it complex with many interrelated components or was it simple and straightforward?

	1	2	3	4	5	6	7	
Simple	○	○	○	○	○	○	○	Complex

How many variables were changing within the situation? Were there a large number of factors or were there few changing variables?

	1	2	3	4	5	6	7	
Few	○	○	○	○	○	○	○	Large

How alert were you during the situation? Were you alert and ready for activity or did you have a low degree of alertness?

	1	2	3	4	5	6	7	
Not alert	○	○	○	○	○	○	○	Alert

How much did you have to concentrate on the situation? Were you experiencing full concentration or little to no concentration at all?

	1	2	3	4	5	6	7	
Little	○	○	○	○	○	○	○	Full

How much was your attention divided in the situation? Were you focusing on many aspects of the situation or focused on only one?

	1	2	3	4	5	6	7	
One	○	○	○	○	○	○	○	Many

How much spare mental capacity did you have during the situation? Did you have enough to attend more variables or nothing to spare at all?

	1	2	3	4	5	6	7	
None to spare	○	○	○	○	○	○	○	Enough to spare

How much information were you able to collect about the situation? Did you receive a lot of information or very little?

	1	2	3	4	5	6	7	
Little	○	○	○	○	○	○	○	A lot

How much did the received information help with understanding the situation? Have you understood a great deal of knowledge or very little?

	1	2	3	4	5	6	7	
Little	○	○	○	○	○	○	○	A lot

How familiar were you with the situation? Did the situation feel familiar (similar to past ATC experiences) or was it a completely new situation (unlike any ATC experience you have had) ?

	1	2	3	4	5	6	7	
Not familiar	○	○	○	○	○	○	○	Familiar

Did you feel bored during the session?

	1	2	3	4	5	6	7	
Not bored at all	○	○	○	○	○	○	○	Very bored

Please rate the complexity of the traffic during the experiment. Did you find the scenario overall simple (easy) or complex (difficult)?

	1	2	3	4	5	6	7	
Simple	○	○	○	○	○	○	○	Complex

What was the overall workload level you experienced during the experiment?

	1	2	3	4	5	6	7	
Very Low	○	○	○	○	○	○	○	Very High

Please explain your previous answer and elaborate more on the workload level you experienced. Was it overwhelming, underwhelming, balanced, or was it variable over time?

What strategies did you use during the experiment, how did you approach the tasks? For example, eye scanning patterns, points of focus, etc. Please elaborate.

What was your order of priorities during the experiment? What did you focus on more and what less (waypoints, real aircraft, fictional aircraft, restricted areas, etc.)? Please elaborate.

Did you feel engaged during the experiment? Were you vigilant during the whole session? Do you feel the presence of fictional aircraft helped? Please elaborate.

How would you rate the automation use and actions during the experiment? Was it a positive or a negative experience?

	1	2	3	4	5	6	7	
Negative	○	○	○	○	○	○	○	Positive

Did you find automation to be predictable and understandable? Was the feedback from automation adequate? Please elaborate.

What was your overall level of trust in automation throughout the experiment? Was it high or low?

	1	2	3	4	5	6	7	
Low trust	○	○	○	○	○	○	○	High trust

How did your trust in automation change throughout the experiment? How was it affected by anomalies? Please elaborate.

Did the presence of automation anomalies make you feel more vigilant? Did your vigilance fade over time? Please elaborate.

How prepared did you feel for taking over manual control after the automation failed?

	1	2	3	4	5	6	7	
Unprepared	○	○	○	○	○	○	○	Prepared

Please explain and elaborate on your previous answer. Why did you feel prepared or unprepared?

What is your opinion on the number of fictional aircraft present on screen? Was it too little, too much or balanced?

	1	2	3	4	5	6	7	
Too little	○	○	○	○	○	○	○	Too much

Please explain and elaborate on your previous answer.

How did you find the distribution of fictional aircraft in the sector to be? Were they well or badly distributed?

	1	2	3	4	5	6	7	
Badly distributed	○	○	○	○	○	○	○	Well distributed

Please explain and elaborate on your previous answer.

How did you see and treat fictional aircraft? Were they more of a tool/game element or did you see them the same as actual traffic? Please elaborate.

Do you feel like fictional aircraft helped you in remaining vigilant and engaged? Please elaborate.

What is your opinion on the overall implementation of fictional aircraft?

Do you have any other comments?

F

Results Summary

The following appendix presents raw data collected during the experiment. Odd numbered participants were in the fictional aircraft group and even numbered participants were in the baseline group.

F.1. Selected Data Plots

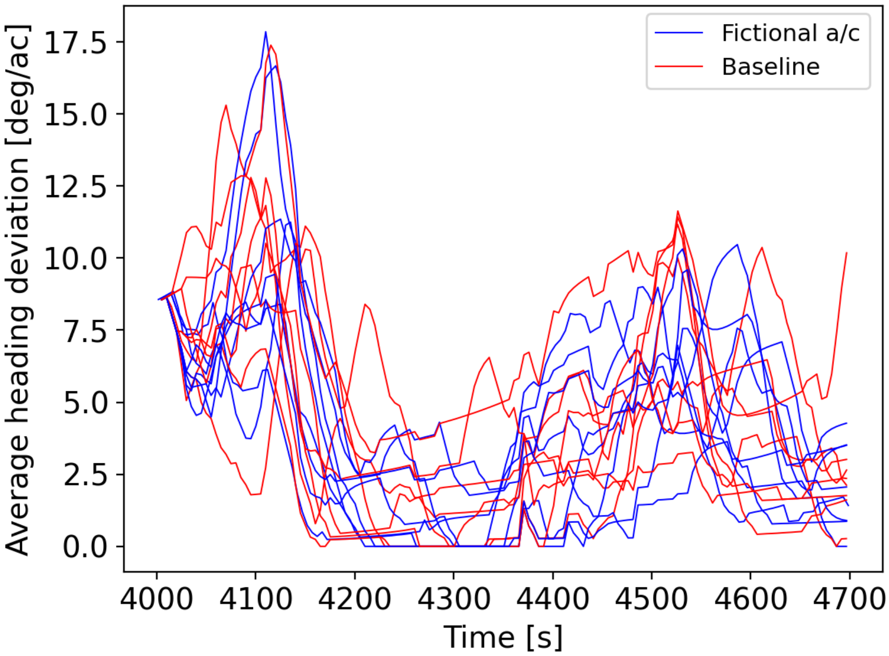


Figure F.1: Heading error indicator in function of time after the failure event for all participants.

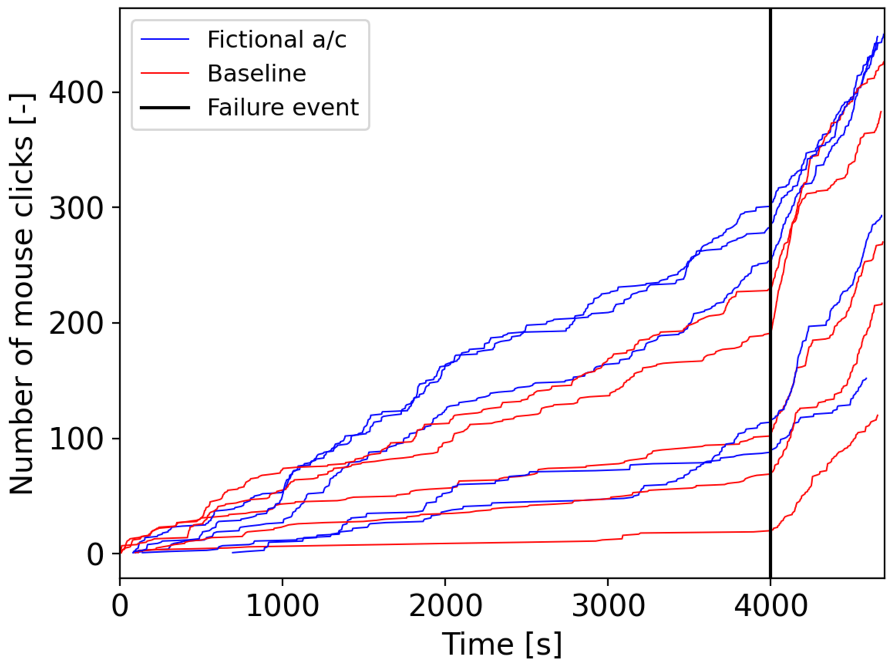


Figure F.2: Cumulative number of mouse clicks over time for participants 7 to 16.

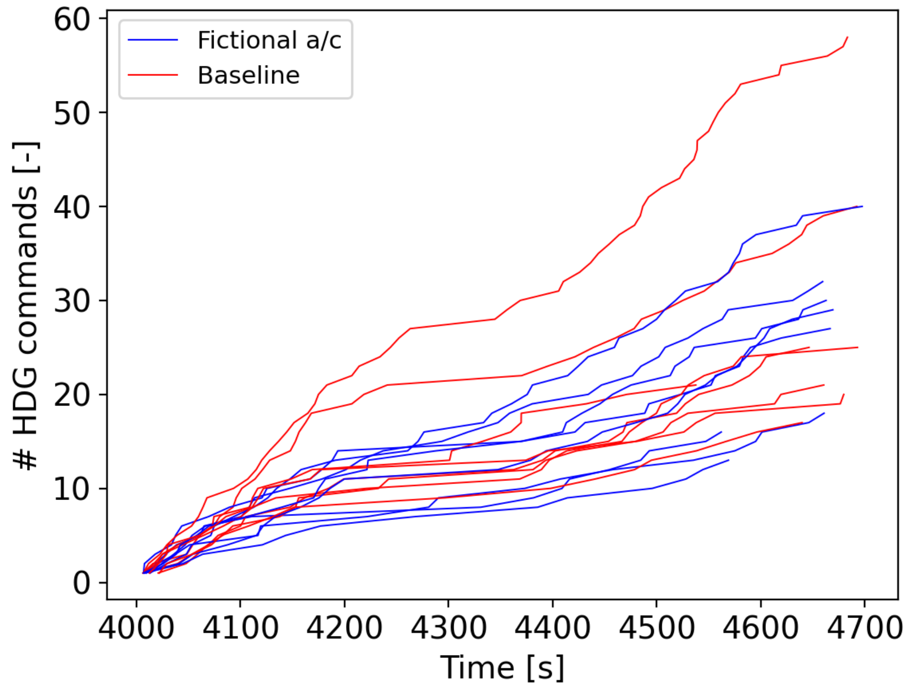


Figure F.3: Cumulative number of heading commands over time for all participants.

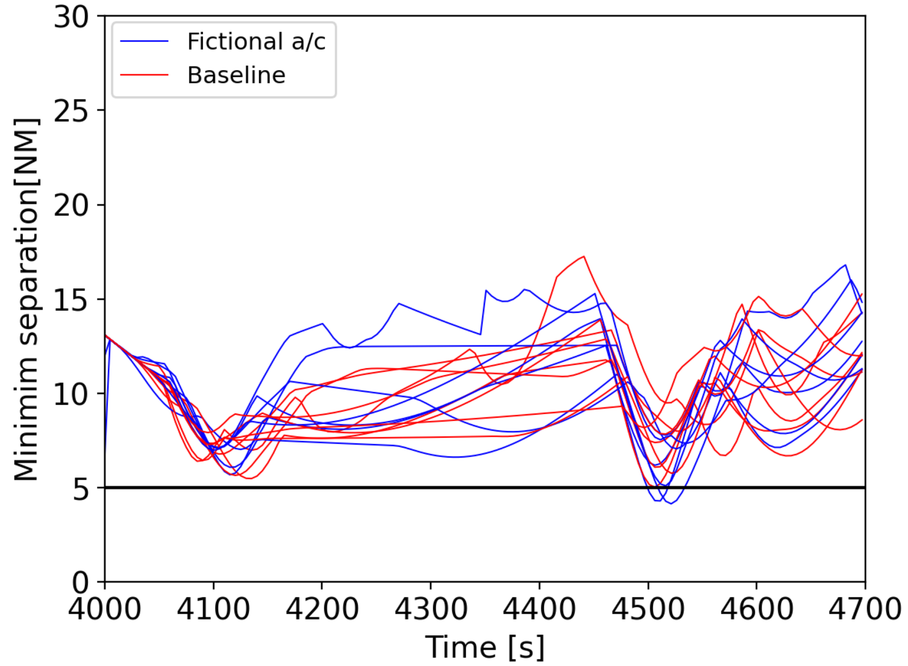


Figure F.4: Minimum separation over time for all participants.

F.2. Performance Data

Table F.1: Summary of recorded parameters for all participants.

Subject number	Additional track miles [NM]	Failure alert reaction time [s]	Number of real a/c HDG cmds.	Number of fic. a/c HDG cmds.	Total a/c clicks, both types	Label clicks before failure	Label clicks after failure	Global click rate ratio	Label click rate ratio
1	38.28	3.05	16	12	53	130	30	3.04	1.32
2	33.23	1.54	37	0	68	31	12	14.75	2.21
3	33.43	2.27	25	30	100	34	8	8.94	1.34
4	34.75	9.20	56	0	129	12	5	63.81	2.38
5	37.91	2.71	32	37	63	64	14	3.41	1.25
6	36.40	1.25	24	0	82	52	8	9.89	0.88
7	36.89	1.43	31	38	71	75	7	3.51	0.53
8	35.68	2.74	21	0	53	13	8	26.81	3.52
9	42.96	2.17	13	18	19	24	4	3.06	0.95
10	35.87	1.82	25	0	53	90	33	5.46	2.10
11	36.90	2.45	16	23	50	76	6	2.50	0.45
12	39.67	1.54	21	0	83	168	76	5.41	2.59
13	36.66	2.25	26	26	65	20	4	8.39	1.14
14	34.86	3.20	16	0	46	60	11	5.43	1.05
15	34.23	1.42	28	29	45	125	20	2.32	0.91
16	52.89	4.40	20	0	38	111	25	3.24	1.29

F.3. Anomaly Reporting Data

Table F.2: Time of each anomaly report for each participant, as well as number of false positive reports and number of anomaly report window closing without reporting.

Subject number	Anomaly 1 [s]	Anomaly 2 [s]	Anomaly 3 [s]	Anomaly 4 [s]	Anomaly 5 [s]	Anomaly 6 [s]	Anomaly 7 [s]	False positive reports	Close without reporting
1	788.788	1644.931	2037.134	2497.009	2969.585	3085.011	3897.038	3	3
2	943.322	1640.015	2025.367	2485.958	2927.75	3247.161	3838.106	0	3
3	884.381	1804.097	2039.367	2509.779	2961.026	3226.741	3885.53	0	0
4	788.267	1773.065	2036.05	2506.61	2962.536	3182.933	3863.531	0	3
5	843.4	1659.001	2034.353	2537.157	2948.468	3191.844	3862.146	0	4
6	792.061	1643.619	2037.052	2485.662	2952.087	3165.262	3834.94	0	0
7	793.433	1668.252	2030.172	2447.192	2966.907	3247.755	3842.896	2	7
8	790.422	1626.661	2019.251	2494.552	2924.552	3188.848	3894.172	0	0
9	794.541	1633.822	2040.14	2499.234	2916.456	3097.417	3833.032	0	2
10	807.239	1788.169	2029.501	2503.804	2955.675	3244.759	3866.779	0	6
11	762.326	1613.744	2001.436	2483.228	2889.031	3001.46	3815.064	2	7
12	773.309	1614.415	2026.334	2438.409	2788.505	3076.309	3776.579	0	5
13	780.142	1803.176	2024.493	2489.187	2921.822	3275.678	3836.234	1	0
14	784.604	1629.875	2040.499	2504.585	2952.914	3251.685	3883.97	0	3
15	787.115	1762.085	2016.756	2474.492	2940.854	3011.444	3823.925	1	0
16	789.33	1645.522	2031.248	2235.312	2845.32	3158.35	3518.695	1	7

F.4. Answers to Survey Questions

F.4.1. Questions Answered by Both Groups

How changeable was the situation? Was it highly unstable (1) and likely to change suddenly or stable and straightforward (7)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	6	6	5	6	6	5	6	5	5	5	6	6	5	6	6	4

How complicated was the situation? Was it complex with many interrelated components (7) or was it simple and straightforward (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	2	5	5	2	3	3	4	2	2	4	2	3	2	2	3	2

How many variables were changing within the situation? Were there a large number of factors (7) or were there few changing variables (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	2	2	3	4	2	2	3	2	2	2	2	2	2	2	4	2

How alert were you during the situation? Were you alert and ready for activity (7) or did you have a low degree of alertness (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	3	3	4	4	5	5	6	5	6	6	4	3	4	6	5	6

How much did you have to concentrate on the situation? Were you experiencing full concentration (7) or little to no concentration at all (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	4	5	3	2	6	4	6	3	5	5	6	2	5	3	5	6

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)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	4	6	5	5	5	6	7	3	5	3	6	6	4	4	6	5

How much spare mental capacity did you have during the situation? Did you have enough to attend more variables (7) or nothing to spare at all (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	6	4	6	6	6	6	4	7	6	7	6	6	6	7	5	3

How much information were you able to collect about the situation? Did you receive a lot of information (7) or very little (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	5	4	5	6	4	6	5	6	6	2	2	5	6	5	4	6

How much did the received information help with understanding the situation? Have you understood a great deal of knowledge (7) or very little (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	3	6	6	5	5	6	6	7	5	6	2	5	5	5	5	6

How familiar were you with the situation? Did the situation feel familiar, similar to past ATC experiences (7) or was it a completely new situation, unlike any ATC experience you have had (1)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	6	5	7	3	6	5	2	7	6	7	3	6	7	6	6	5

Did you feel bored during the session (1 - not bored, 7 - very bored)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	7	6	6	5	5	6	1	7	5	6	6	6	5	5	5	5

Please rate the complexity of the traffic during the experiment. Did you find the scenario overall simple (1) or complex (7)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	4	3	3	2	3	3	5	2	2	4	2	4	4	5	3	2

What was the overall workload level you experienced during the experiment (1 - very low, 7 - very high

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	2	2	3	3	3	3	5	3	2	1	6	2	3	2	3	4

Please explain your previous answer and elaborate more on the workload level you experienced. Was it overwhelming, underwhelming, balanced, or was it variable over time?

- **Participant 1**
In the manual part the workload was nicely balanced. In the automated part my workload was way too low to keep awake. There were occasional peaks when I had to give a control to a (fake) aircraft, or when reporing an anomaly, but the monitoring task required very litle workload

•Participant 2
In automation mode, the workload was very low. The only thing to do was to check for anomalies. Therefore, I started thinking about other suff (that I sill need to do today;)) and started to moan, play drums with my fingers etc etc. However, when automation failed the workload was suddenly very high. Multiple conflicts needed to be solved right away. In the end I let 1 AC fly in restricted airspace so that is not so good.

•Participant 3
The workload was quite low, most aircraft followed the correct route and the virtual aircraft didn't require much attention

•Participant 4
Underwhelming throughout most of the experiment, as you learn to watch out for certain specific areas for the automation mistakes (High traffic areas for LOS, end of trajectory for wrong waypoint and the restricted areas for the restricted breaches) Which means that overall you are only paying atention to certain areas of traffic when automation is on because you know that for the others its not a problem. Once automation is off workload increases significantly but is still manageable

•Participant 5
The overall workload was quite low, the automation managed to solve 95% of cases without any anomalies. The perceived workload and complexity to route the virtual aircraft was also quite low. When the automation failed a real spike was felt in the workload and complexity though. It took quite some metal effort to route all remaining aircraft without generating any conflicts.

•Participant 6
The fact that I felt bored during the session, is, I believe, one of the things that "showed" that my workload level was low. Also, the complexity of the traffic was fairly low, as the aircraft did not differ much in their airspeed and many aircraft headed in the same direction(s), as there were a few fixed air routes. I felt that my overall workload was pretty low, as I felt that it was not too difficult to monitor the aircraft, as the update rate of the screen was low and the air routes made it fairly simple to monitor whether the aircraft were going in the right direction. Also, when the automation failed and I had to take over, not too many aircraft were involved in a conflict at the same time, so the conflicts weren't too difficult to solve. I do believe that two protection zones touched once when I had taken over from the automation, but other than that, I thought it didn't go too bad. The fact that I was already familiar with the 'rule of thumb' of steering the slower aircraft behind the faster aircraft before taking part in this experiment made it easier for me, and thus lowering my workload. And also because I have taken part in a few ATC experiments in which the conflicts were way more complicated to solve (e.g., more aircraft involved in a single conflict), made it feel to me like my workload was rather low in this experiment.

•Participant 7
During the automation there was little work load, other than learning and checking the automation. However, once the automation suddenly turned off, there was a lot of work trying to

sort out the automation resulting in close calls. So the work load came with spikes. However, once sorted, it was easy enough.

•Participant 8
Rather low; most of the time nothing to do

•Participant 9
Mostly low workload, only slightly higher when situations with anticipated required action occurred

•Participant 10
Very underwhelming. I really did not have to do too much. It changed when there were moments near a conflict, but those had usually been anticipated so it did not result in very high workload. The final manual mode of course required a higher workload but also not that much.

•Participant 11
The workload was quite high, but with (in my view) unnecessary actions for a relatively simple scenario. For example, when checking if an aircraft is routed towards the correct waypoint, the aircraft heading needs to be monitored after a command has been issued by automation. It would make more sense to me if I was informed on what decision was made and why. This would increase my understanding of the automation, likely lower my workload and shorten required time to recognize any anomalies.

•Participant 12
Workload level felt very low, it was underwhelming. There were only a few automation anomalies in the end, and it was quite boring to only watch and verify that the automation acted as expected.

•Participant 13
workload variable over time. while supervising the automation, the workload was low with peaks of (slightly) higher workload, especially when I saw a situation that might require me to report a malfunctioning of the automation, and the fictitious aircraft were distracting me from it. when the automation failed, workload was average.

•Participant 14
It wasn't much work. The scenario did contain some situations where two things needed attention at the same time, but not so much that they couldn't be handled. The workload was acceptable, with a rate of failures that was high enough to warrant constant attention.

•Participant 15
I think especially the monitoring (including the virtual AC) was quite underwhelming, with a couple of peaks when multiple events happened at once. I don't think I necessarily missed anything during these peaks, but was maybe a bit late to note some anomalies.

•Participant 16
What used a lot of my workload were situations (possible future conflicts/ flying in restricted area's, wrong heading to waypoint) which were not solved yet, because I believe that when I detect a future conflict, the automation is already too late. Because what I've learned is that

whenever you detect a problem, you immediately solve it. And so I have to keep monitoring (demands high workload), when the automation will solve the problem, which I would've already solved.

What strategies did you use during the experiment, how did you approach the tasks? For example, eye scanning patterns, points of focus, etc. Please elaborate.

•Participant 1
At some point I started using the label positions to indicate whether an aircraft was on course or needed to be steered to some direction (in which I would place the label). If I spotted a possible conflict, I would keep going back to it during my scans, until I knew i wouldn't turn into a conflict. Every now and then I looked at the restricted areas and searched for any aircraft that were on course to get near them.

•Participant 2
During automation mode I had only to supervise the automation and check for anomalies. There were 3 type of anomalies so I subsequently looked for the different types. So first check for AC leaving the airspace and check whether they are leaving at the correct exit point. Then check for AC that were about to or were in restricted airspace, finally check for AC close to each other for Loss of Seperation (and in case automation failes) and then repeat these steps. When automation failed I started with solving conflicts and then solve exit way points and restricted airspace.

•Participant 3
Trying to predict where a problem could occur and focus on these aircraft. On a certain point it was either resolved or not, in both situations I could continue to focus on other aircraft (after I had sent a report about the anomaly in the second siuation)

•Participant 4
For the supervisory task I tried to understand the overall strategy of the automation (if it used direct routing or how it solved possible conflicts) and then tried to monitor for where I could see errors ocurring (next to restricted areas, high traffic density areas or at the end of trajectories). A regular eye scanning to see which planes were coming in and if there were any planes I had not payed attention to in a while. Aside from that, during manual control I usually did a regular heading check on each aircraft after solving each conflict. I also tried to extrapolate where the plane trajectories would intersect and estimate if they would reach that intersection at the same time (possible conflict)

•Participant 5
In general I used a scanning pattern, roughly going clockwise around the outside of sector to check if in- and out-coming aircraft are headed towards the correct waypoints and if the 2.5 NM separation circle was going to interfere with restricted airspace. After the circular scan of the outsides of the sector a more detailed scan was carried out of all remaining aircraft to see if they were not going to experience a LOS.

•Participant 6
During the monitoring, I just kept going over the aircraft one-by-one with my eyes in a repeat-

ing pattern to check whether they were deviating from their intended track or not and whether I could foresee a conflict. And when I saw an aircraft label turning yellow, which indicated that the respective aircraft was given a heading command by the automation, I paid attention to that particular aircraft for a little while to see if it was steered in the right direction or whether I could foresee an anomaly. When the aircraft was steered on the correct air route, or at least in the right direction, I continued my repeating pattern of going over the aircraft. Also, I checked the intended exit waypoint of aircraft when I saw they were entering the sector to see whether they were already going in the right direction, or whether they would still have to be steered by the automation in the right direction. If I felt I had sufficient capacity left, I also checked the aircraft that were still outside the sector, to already see what their intended exit waypoints were. After the automation had failed, and I had to take over, I did a lot of clicking on the aircraft, to check whether they were heading towards their intended exit waypoints or whether I had to give a heading command to steer them in the right direction. Also, when I saw that a conflict was coming up, I tried to obey the ”rule of thumb” of steering the slower aircraft behind the faster aircraft.

•Participant 7
I learned the relative dimensions and speeds from the automation. From that I learned how close I can get aircraft. I tried to let them follow direct routes and push the limits of the separation as much as possible and predict from as far away as possible.

•Participant 8
check exit for every new aircraft; then check direct and look for potential conflicts; if necessary, focus on intersection points to see if actual conflict occurred

•Participant 9
Cyclic scan over aircraft, with additional attention for aircraft that will need an action at some point. Looked at bearing between aircraft for pairs that had the potential of getting too close.

•Participant 10
When there were no imminent conflicts, I would carefully check the trajectory of each aircraft. If I felt like they were heading in the right direction and no conflict was imminent I moved the label to the bottom and away of the aircraft point (so I could recognize the aircraft that were doing well in an instant). For aircraft that needed a heading change or those close of a loss of separation (both with other aircraft and with no fly zone), I moved the label pretty much on top of the aircraft. This made me quickly see which aircraft needed closer attention and more frequent checking. When there was a conflict very close by, I would do my best and NOT focus only on that conflict but actually continue to scan the whole environment for any other conflict. As soon as something happened I would take note with the text box and move on.

•Participant 11
Constantly scanning the sector in a top to bottom fashion, identifying potential hazards and keeping a close eye on them.

•Participant 12
I used the position of labels as an indication of what automation behavior I expected. For instance, aircraft on a route that split into a north/south stream, I would place the labels of

the aircraft planned to go south, south of the aircraft symbols and vice versa for north. That made it easier to check if the automation steered them towards the correct route. Further, if aircraft were already heading direct to the planned exit point, I would put the label on top of the aircraft symbol. In case the automation made a change to that aircraft, it was straightforward to check if the aircraft veered off course or not. Further I kept a constant scan pattern for aircraft not flying on a ”fixed” route to monitor if they interfered with other traffic, would breach a restricted area, or cause other issues.

•Participant 13
While supervising the automation, i scanned mostly following the highways lines. from bottom to top. i kept count of how many situations might have required my attention in the near future, re-evaluating the situation when a new aircraft came in the airspace. Once I found a situation that would require my immediate attention, i checked none of the other interactions were as urgent, and spent some time focusing on the alarming situations

•Participant 14
First check if each aircraft is on track, and focus on aircraft that have a decision point coming up. Next, look for potential conflicts (aircraft and forbidden zones) and predict solutions. Monitor if solutions are being implemented by the automation.

•Participant 15
I put the labels on the back of all the AC and if I thought that a conflict might occur I would put the label far away from the AC, then I would usually scan bottom to top and keep checking the destinations. If I noticed a possible conflict I would scan it more than the others which I thought/knew were fine. However this sometimes led to me being too focused on one event I think, which meant that I was a bit late to notice some others.

•Participant 16
I put the label over the aircraft which had no more conflicts, and was heading towards its destination. Therefore the aircraft’s behind their own label I wouldn’t be needed to give any attention to (lowering my workload). I chose to prioritize my workload (safety), rather than flying more efficient, meaning I choose to give an aircraft a (larger) heading which I was certain would solve a conflict, than rather give a smaller (more efficient) heading, which would require me ”monitoring” and demand a higher workload.

What was your order of priorities during the experiment? What did you focus on more and what less (waypoints, real aircraft, fictional aircraft, restricted areas, etc.)? Please elaborate.

•Participant 1
It took me a while to spot one fake aircraft (coming from the south), so I think I was mostly focused on the real aircraft. Possibly because their colours stood out more against the background? Restricted areas I scanned occasionally. Waypoints I only looked at when I found a label with a waypoint which location I didn’t remember.

•Participant 2
In automation mode, I was focused more on the waypoints, because it is harder to spot

an error here. AC collision and AC close to restricted airspace are easer to spot. During automation fail mode, I focused more on the aircraft and subsequently on the restricted areas and waypoints

•Participant 3
From high to low: waypoints, restricted areas, los, fictional aircraft. The restricted area and los anomalies were easier to detect, since it was clearly visible due to the circle(s) overlapping, the waypoints actually required you to keep track of which waypoint was where on the map. The fictional aircraft even stood out more due to their colour and where easy to just observe sometimes to see if they didn't get into trouble.

•Participant 4
Most important to least important would be Aircraft-Waypoint-Restricted area. I barely looked at the restricted area if there was no aircraft around. The waypoints I only payed attention to when monitoring automation (for the mid point) or to make sure they were heading in the right general direction when I made corrections. Aircraft I kept my eye on to estimate overall traffic density and flow. I also payed closer attention to fast aircraft and those that were coming coming from above or bellow (as most traffic was going left to right or vice versa, these aircraft that had a vertical path would intersect the highest amount of flight paths, hence a higher number of conflicts)

•Participant 5
I tried to focus more on real aircraft, continuously checking if they were headed towards the correct waypoints and if they were possibly going to encounter a LOS in the near future. Restricted areas also had a high priority, but less focused on since they were only present at the edges of the sector.

•Participant 6
I think I relied a lot on the air routes visible during the automation monitoring. I looked a lot at the aircraft labels to check whether aircraft were heading in the right direction. Also, when I saw that an aircraft was entering the sector close to a restricted airspace or was required to leave the sector close to a restricted airspace, at those times I tried to pay a little extra attention on whether those aircraft actually invaded the restricted airspace. Sometimes when an aircraft was entering the sector, the yellow label made it clear in a very easy manner that the aircraft was given a heading command to avoid entering the restricted airspace. I do believe I unconsciously paid more attention to the intended exit waypoints than on avoiding losses of separation, perhaps due to the presence of these visible air routes during automation monitoring. This is of course not necessarily desirable, as safety is priority number one, so making sure that no losses of separation occur should be the main priority both during monitoring and during ”manual” control.

•Participant 7
In order; real aircraft making mistakes with separation, restricted areas, way points and then the virtual aircraft in the same sequence.

•Participant 8
All equal; more issues with waypoints and aircraft, hence also more focus on those. RAs

were easily spotted

•Participant 9

First detected potential conflicts, then real aircraft with required actions, then other aircraft and fictional aircraft. safety first.

•Participant 10

Aircraft loss of separation needed most attention, because the other two could be foreseen a long time before it took place. So if a wrong heading or intersect with no fly zone took place, it was not a surprise usually.

•Participant 11

Firstly, focus on real aircraft. Order of priority: 1) conflicts 2) restricted areas 3) correct waypoint 4) fictional aircraft

•Participant 12

I focused the most on keeping track whether the traffic stream conformed to the planned routes by the above mentioned label-placement strategy. Next to that I focused on non-standard aircraft (ones not flying along a fixed route) to monitor their conformance, restricted area avoidance, and conflict avoidance.

•Participant 13

I kept track of how many aircraft would possibly require my attention, indiscriminately from the type of conflict/problem that might arise. When I saw an immediate problem I would focus on that, again indiscriminately (ac too close to the restricted zone, or to another aircraft, or too close to the wrong exit point). I kept the fictional aircraft in the back of my mind, I adjusted their course in correspondence with the waypoints inside the sector, and last minute in case of conflict. Prioritizing the real situation over the fictitious one, caused me to be too late in avoiding conflicts between the real and fictitious aircraft, which then caused me to loose focus on the real aircraft while trying to solve the conflicts I generated with a fictitious aircraft.

•Participant 14

See previous answer.

•Participant 15

I think it was easiest to focus on the restricted areas, since they were on the edge and there wasn't a lot of traffic around them, so that didn't take that much effort. Then I think I spent most of the time looking for possible real aircraft conflicts and thinking about what I would do to solve them. This sometimes took more time than I expected I think because the automation sometimes cut it quite close, which meant I would be monitoring a situation that in the end was fine, meaning I might have missed something else. In terms of the Waypoints, I think that it was quite easy. The AC take a lot of time to move across the control area so you had a lot of time to check and double/triple check. Although again sometimes you wouldn't be sure if the automation was just making a detour or actually sending it to the wrong waypoint. It would be nice if the automation showed the intended final waypoint, but I guess that defeats the purpose of this task.

•Participant 16

I had enough time (and workload) to focus on all 3, but of course if I had to make a chose between the 3, it would be loss of separation.

Did you feel engaged during the experiment? Were you vigilant during the whole session? Do you feel the presence of fictional aircraft helped? Please elaborate.

- **Participant 1**
My vigilance was ok at the start and then dropped dramatically. I go distracted by looking around me, had to yawn a lot and generally felt very bored :-) The fictional aircraft helped a little bit but I' d have liked more of them to really keep me awake. The manual part was much more fun, rewarding and exciting.
- **Participant 2**
No I did not feel engaged during the whole experiment. I had to put a lot of effort in to stay focussed. Presence of fictional aircraft would have helped to stay more engaged but I am not sure if I would have found all anomalies
- **Participant 3**
I did notice losing concentration some times, since not a lot of things happen. It was quite an easy task, and the anomalies where generally easily detected. The fictional aircraft were to easy to direct in this scenario so I am a bit skeptical about the usefulness of their presence, I can imagine that with more fictional aircraft a higher concentration might be required.
- **Participant 4**
I felt engaged, I did feel like during the automated stage I delegated a lot to the automation and after certain criteria were met for each aircraft (no obvious close encounters or flight path intersection and they were heading in the right general direction without going straight for a restricted area) my atention for each aircraft and in general tended to drop. I believe fictional aircraft would have helped with countering the boredom and the complacency that starts settling in after a while. The traffic patterns become fairly clear and you start to lose attention
- **Participant 5**
I feel like the presence of the fictional aircraft really helped me stay vigilant and to combat getting bored... I did find that the amount of times I had to take action to keep the virtual aircraft headed towards the correct waypoint and separated from restricted airspace and other aircraft was quite low, which resulted in slightly less vigilance even when virtual aircraft were in screen. This is of course a difficult balance because one can start to focus too much on the fictional aircraft which can result in anomalies or even failures of the automated system controlling the real aircraft to go unnoticed.
- **Participant 6**
I did not feel engaged that much. This resulted in me feeling bored, and sometimes perhaps loosing a little bit of attention. I do believe the presence of fictional aircraft might have helped, because if I would have had to give actual commands and do some more clicking, rather than only staring at the screen, I believe I might have stayed a little bit more alert. However, the downside might have been that I would have paid a little less attention to the ”real” aircraft,

but I think it would be possible to avoid that by not generating too many fictional aircraft, just enough that some clicking and giving heading commands would be necessary every once in a while, to stay alert.

•Participant 7

I was vigilant during the whole thing, I think that for the duration of the experiment it did not help me stay focused, but I think that for longer periods of time it will definitely help. The virtual aircraft also let me test some theories and patterns. Experimenting as a whole was very nice and good to keep up skills.

•Participant 8

It would have been nice to have something to do. Not necessarily aircraft, just something else to not be so bored (e.g. talk or drink coffee)

•Participant 9

I tried to stay alert, but the lack of interaction was sometimes affecting my vigilance. I did not have the impression that the fictional aircraft helped.

•Participant 10

I would continue to scan but I was not very engaged. I actually don't think fictional aircraft would have helped since the workload was so extremely low. I think if it were slightly more chaotic then the use of additional aircraft would help me focus. But now it was 'too easy' and therefore not a problem to notice any conflict. (at least I hope I didn't miss any)

•Participant 11

I felt more engaged when I had full control (i.e., after the automation was turned off). The presence of the fictional aircraft helped somewhat, but the fact that you know that the aircraft is fictional makes you less alert. I sometimes had the feeling of 'what's the point in controlling these aircraft if they don't even exist'.

•Participant 12

I did not feel engaged during the experiment with automation engaged. Only when the automation failed, I instantly became more alert and started thinking more pro-actively. When only monitoring automation I noticed that I had some issues with vigilance. I think that performing a secondary control task with fictional aircraft would have increased both my awareness and alertness.

•Participant 13

I was vigilant, but not engaged during the part in which i had to supervise the automation. I feel like the fictional aircraft only were in the way of my general comprehension of the situation. I did not feel more engaged by having to sporadically control those. I found myself bored, but i never lost focus.

•Participant 14

I did not feel I was getting distracted. The level of traffic was high enough to avoid being bored and distracted. There was always something to look at. With fictional aircraft the interactivity would be higher, but in this scenario I don't think it would have prevented loss of engagement.

•Participant 15

I think there were short moments where I sort of wondered off with my thoughts, but in general

I don't think this impacted my vigilance since it was only during moments where very little possible conflicts seemed to happen. I do think the fictional aircraft helped to stay vigilant, although I also think they sometimes were a little distracting.

•Participant 16
The workload during the experiment was not too much so some parts were boring, however it was still stressful because I did not feel in control, and situations happened which I would have solved differently. About the fictional aircraft, no, I have a certain amount of attention I can give, I and would rather give 100% of my attention to real aircraft.

How would you rate the automation use and actions during the experiment? Was it a positive (7) or a negative (1) experience?

Participant No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Answer (1-7)	6	3	6	6	6	6	7	6	5	3	2	4	4	5	5	2

Did you find automation to be predictable and understandable? Was the feedback from automation adequate? Please elaborate.

•Participant 1
It was quite predictable, as I often found looking at an aircraft, thinking it should get a command and then a second later the computer indeed gave a command. I would've liked to know what the actual commands were, so I didn't have to wait for aircraft to actually start turning to see which direction automation sent them.

•Participant 2
Yes the automation was predictable, even the errors the automation made were predictable

•Participant 3
Most of he times yes, but sometimes the computer decided to put the aircraft in a certain direction which I wouldn't have picked. Eventually it worked out most of the times. The anomalies only seemed to occur when the computer didn't do anything.

•Participant 4
I found it to be more agressive in its solutions than I would have liked. Sometimes the automation would make a conflict resolution that to me seemed like it would still lead to an LOS or breach of restricted space but would then turn out to be okay. My main gripe with it was that there was no feedback on the automation part on whether it was detecting a possible conflict until it actually did a correction (which meant sometimes I was looking at a pair of aircraft or at an aircraft going near a restricted area for a while trying to figure out if the automation had detected that and would correct for it or not)

•Participant 5
The automation was in general quite predictable. It would have been handy though to see an orange line showing the new commanded heading when the automation was giving instructions to aircraft. There were a few situations were I couldn't anticipate the heading change which resulted in a higher workload to route the virtual aircraft.

•Participant 6
I do believe the automation was very predictable due to the air routes being visible during monitoring. I thought that was really nice. I also liked the fact that when aircraft were given a command by the automation, the aircraft label turned yellow, because that was really noticeable, so then I could pay attention for a short amount of time on what the automation was changing. The airroutes also made it a little easier to predict when an aircraft was heading towards the wrong exit waypoint.

•Participant 7
It was very predictable and understandable. I would like the automation to explain more as to what it is doing. Sometimes it was hard to distinguish avoidance and heading mistakes.

•Participant 8
Yes, it was predictable, nothing unforeseen happened

•Participant 9
It would be helpful to see the cause of actions, i.e., see a STCA alert connected to an automation-initiated manoeuvre.

•Participant 10
Most of the time it was understandable, but a bit slow to respond. I think it would be better if there would be a little alert to say that a heading change will take place but not yet. Because sometimes I would simply wonder when the aircraft would change heading or if it wouldn't at all. So if I could see that the automation program is planning to make a heading change but not yet, then I could relax and trust that it will do so. (maybe turn the aircraft green if all conflicts have been solved and just needs to fly straight)

•Participant 11
The feedback was inadequate, because I had no insight into how and why decisions were made (see also earlier comments).

•Participant 12
The automation mostly conformed to the strategies that I myself would implement. However, how the automation worked, and what the automation based decisions on was completely opaque (black-box). Further, there was no feedback what the automation was doing other than monitoring the "raw" aircraft states, no issued steering cues were provided. The feedback from the automation was thus inadequate to keep me in the loop of what it was doing.

•Participant 13
automation was predictable and understandable. it sometimes causes concern because it acts too optimally for me to distinguish if the action was solving the safety concerns or not. Overall, as a controller I wouldn't appreciate its tuning, while of course I see and understand why the decisions are taken in the points their are, and as a different stakeholder I would appreciate them

•Participant 14
Somewhat predictable and understandable. A direct feedback of the commanded heading would have helped. Now I had to wait to see where the aircraft was actually going.

•Participant 15

As mentioned on a previous question, I think the automation should show the intended final destination waypoint, since this was a lot of guessing. Sometimes I'd wonder why it was making a certain move, thinking it was going to send it to the wrong waypoint, but then it would end up just avoiding a far away conflict with another AC. Overall though I think it was decent, there weren't a lot of times where I had no clue what the automation was doing.

•Participant 16

No it it was more unpredictable than it was predictable. Again I would solve a problem right away, and if the automation doesn't do that, I get stressed and start monitoring and keep watching when the automation is going to start doing something. Also when automation does something, it takes time for me to understand why it is doing something, which I would rather not spend my metal load on, I wound my workload only to consists of keeping it safe, and not trying to understand what the automation is doing.

What was your overall level of trust in automation throughout the experiment? Was it high (7) or low (1)?

Participant No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Answer (1-7)	4	2	5	5	6	6	5	5	6	6	2	2	5	5	5	2

How did your trust in automation change throughout the experiment? How was it affected by anomalies? Please elaborate.

•Participant 1

After two aircraft that entered the restricted areas I felt like I couldn't trust automation to respect these areas

•Participant 2

With every anomaly my trust in the system degraded, however also trust in myself to find errors of the automation degraded. Therefore I felt less able to supervise.

•Participant 3

I knew at the start that the automation would make some mistakes, so I was a bit more alerted at the beginning. However, only few anomalies occurred so in the end I trusted the automation a bit more and only focused on the critical situations.

•Participant 4

My trust increased overall as the experiment went on. With the restricted airspace anomalies being common that made me expect that the automation would not respect the restricted area boundaries and thus i started to become more cautious whenever an aircraft got near to it. I also started double checking all headings and waypoints to see if the planes were heading in the right direction. Overall I trusted automation the most to avoid LOS rather than the other two anomalies

•Participant 5

In the beginning I was quite cautious and carefully checking every action of the automation because I didn't quite trust it yet. When the experiment progressed there were only a small amount of anomalies which greatly increased the trust in the automation. After an anomaly was detected though my trust in the automation decreased slightly for a small period of time.

•Participant 6
I think at the very start, the trust still had to be built and I was waiting for that alert that the automation had failed. However, because there were very few anomalies, I believe that the trust grew. As I did not expect automation to be perfect due to my study background, the few anomalies that occurred did not really affect the trust that I had in the automation later on in the experiment.

•Participant 7
I found that the restricted airspace breaches had less of an impact compared to the breach of separation. Once I realized that automation failure could happen during separation breaches, I thought it would be very unsafe when imagining real life.

•Participant 8
Anomalies were expected and usually spotted well in advance, hence no change in trust

•Participant 9
Some anomalies were somewhat artificial (going the wrong way while the exit waypoint is correctly present in the aircraft label seems unlikely). I can imagine that it would reduce the trust in automation if this happens regularly.

•Participant 10
It seemed rather predictable, and also when it was doing something wrong I would wait until the loss of separation occurred and immediately take note.

•Participant 11
If you know that the automation can (and does) make mistakes on a safety level, you don't trust the automation anymore and start checking every single decision. This rapidly increases your workload.

•Participant 12
I expected to encounter anomalies, so my trust in automation was low at all times during the experiment. The anomalies themselves did not change that feeling throughout the experiment. Perhaps if the anomalies would have been more unexpected (not briefed before the experiment), this would have been otherwise.

•Participant 13
Most of the control was effectively taken care of by the automation. Once the aircraft had been routed to the correct waypoint, i didn't see any be rerouted again. So I trusted the automation to keep the situation stable after solving a conflict. I did supervise more closely all the points where danger could arise very quickly, such as next to restricted areas of in close vicinity to other aircraft. I am surprised so many anomalies where present in the experiment run, because I have the feeling that it wouldn't be realistic. After all, I find it difficult to believe that ACT could use a software which is unreliable in the most dangerous situations (loss of separation or restricted area invasion). So in the end. it became predictable at what points

in time to expect the automation to not properly correct for anomalies (i'm not talking of the overall failure of the automation)

•Participant 14
Anomalies were expected. The most stressful aspect was the avoidance of aircraft conflicts, because these are more difficult to perceive. This did not change over the course of the experiment.

•Participant 15
I think it got worse over time, again because of the mentioned above. I think the few things that I didn't understand why the automation was doing it, weakened my trust in the automation, and questioning more of its future actions.

•Participant 16
As there were some anomalies by the automation, due to this I felt like I needed to monitor all possible problems (demands workload), and did not trust the automation that it would work without problem.

Did the presence of automation anomalies make you feel more vigilant? Did your vigilance fade over time? Please elaborate.

•Participant 1
Yes, it kept me somewhat awake. But once I noticed that anomalies are not very time-critical I became less vigilant.

•Participant 2
Yes, it woke me up so to say. Finnally something to do

•Participant 3
As explained before, my attention became less due to the little amount of anomalies that occurred.

•Participant 4
I feel like the presence of anomalies made me more cautious about using automation and not necessarily more vigilant. As most possible conflicts were fixed, the few that led to anomalies were too uncommon and far between to make me be continuously vigilant. In order for me to be as vigilant as compared to the manual control these would have to occur more often

•Participant 5
It certainly did. I was constantly on the lookout for anomalies and when I managed to spot and report one it gave a feeling of satisfaction, which increased my vigilance and decreased the level boredom. I noticed that my vigilance did fade slightly over time between the anomalies. One thing I also noticed was when I was filling in the anomaly report I wasn't paying a lot of attention to the rest of the sector. Perhaps shortening the time required to fill in the report by showing a bullet point selection with the most common anomalies would help?

•Participant 6
I do think that the fact that I knew automation anomalies could occur, made me slightly more

alert. However, the anomalies were so rare, that I believe my alertness faded a little bit over time, until I saw a new anomaly occurring. Then my alertness was back for a little bit, after which it faded somewhat again.

•Participant 7
The automation mistakes made me more vigilant and pay more attention to the details.

•Participant 8
it was. the expected anomalies made me actively look for discrepancies which gave good situational awareness

•Participant 9
I specifically looked at situations where an anomaly happened before. This did not change over time.

•Participant 10
No, I think it didn't affect my attention.

•Participant 11
It made me feel more vigilant, but in a wrong way. It felt a bit like double checking a under-trained colleague. I think my vigilance remained more or less constant over time.

•Participant 12
It did make me more vigilant because I expected the automation to fail. But vigilance did fade over time, because only a small number of anomalies occurred.

•Participant 13
It made me feel more vigilant, I didn't lose focus on the task during the experiment run. it might have made me lose the overall picture every now and then while focusing on one single upcoming issue.

•Participant 14
No and no.

•Participant 15
Yes the anomalies were a "fun" break from the monitoring, when an anomaly happened it would definitely make me more vigilant. I don't think my vigilance faded over time, although I do have to say that when the automation failed and I had to take over I still wasn't satisfied with my response time.

•Participant 16
More vigilant yes, but not in a good manner, vigilance only faded over an aircraft when no possible conflicts were on an aircraft route, and the aircraft was heading towards it's destination. If these were not met, if required my monitoring.

How prepared did you feel for taking over manual control after the automation failed (1 - unprepared, 7 - prepared)?

Participant No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Answer (1-7)	6	5	6	5	3	6	6	7	6	6	6	3	7	7	7	7

Please explain and elaborate on your previous answer. Why did you feel prepared or unprepared?

•**Participant 1**
By coincidence I actually looked at an evolving situation that would result in a conflict and tried to come up with a plan to solve it, right when the automation failed. So I could tackle that first problem immediately after taking over.

•**Participant 2**
I saw a lot of conflicts development, so I was aware that automation might be able to handle them and that it thus would fail and hand control over to me.

•**Participant 3**
I did have a good overview of the whole simulation, and already had my eye on a critical situation. So when the automation failed I firstly resolved that situation and then looked at the rest of the aircraft, which were still doing pretty fine.

•**Participant 4**
I felt overall prepared because I had kept an overall idea in mind of how I would redirect traffic when the automation dropped. I felt like some geometric aids such as the lines connecting the middle waypoints to the end ones would have helped in order to keep a better idea of the overall movement but that is a personal preference. I think it would be easy for someone to feel unprepared if they did not know in advance that they would have to take over. Since I knew I would have to take over for the automation failure at some point I kept that in mind. But if I was not sure it would happen maybe that would have affected my preparedeness

•**Participant 5**
I was not quite prepared for the amount of conflicts I had to solve when the automation failed. Maybe partly because the virtual aircraft conflicts were (at least perceived) to be much easier to solve?

•**Participant 6**
I do believe that I was able to stay alert enough. I would not say that I was highly alert all the time, but due to the low update rate of the screen and the number of aircraft present in the sector, I also think that was not necessary for this scenario in order to operate in a safe manner. I feel like it is in my nature to stay focused for quite a long time, so that came in handy in this experiment. However, I do think it was mainly internal motivation to do well that made sure I was prepared to take over manual control. I have a hard time pinpointing something in the experiment scenario that really helped me staying alert.

•**Participant 7**
I had a good overview and idea of the situation. I had routs pre-planned and ways to get out of the mess. However, during the failure, some improvisation was still required considering that the automation had a slightly different setup in the short term that I had to get out of first before stabilizing the situation.

•**Participant 8**
the training was extensive and it was nice to finally have something to do

•Participant 9
I was already sufficiently aware of the situation before the failure occurred, so ready to take over. Only made a wrong estimation (speed difference related) in one of the occurring conflicts after the failure.

•Participant 10
Although I had not yet had a solution in mind, it didn't take long to find one. Also I could see on the edge of the screen that there were no new aircraft on the way so I suspected that the experiment would turn to manual mode soon and end soon. This is a slight design flaw and made me expect the manual control.

•Participant 11
I felt prepared because I had already formed a good mental model of the traffic situation, and thus knew immediately what to do when the automation failed. I do not know if the presence of fictional aircraft played a role in this.

•Participant 12
In the first moments it did come as a little bit of a startle, but when it happened I did already have a picture of where control was required (by using the label strategy). I felt quite prepared to tackle the scenario

•Participant 13
I felt prepared because I knew it could happen, and I was aware of the current situation, and I had already imagined what the automation would do to solve it. I was on the lookout to understand if the automation would behave correctly in that scenario

•Participant 14
The situation was clear at all times. The only surprise was the removal of the "standard route" lines, but that encouraged me to give more "direct to" commands.

•Participant 15
There were times during the experiment where I would be thinking "don't forget, it will happen anytime now" but then when it actually happened I wasn't super fast to response. Although I do think that I was prepared in the sense that I knew exactly what was going on, but maybe not how it would be like in the future, which caused some trouble later down the line.

•Participant 16
Because I was monitoring the aircraft's which needed to change heading in order to solve a possible future problem. The aircraft which no action needed to be taken were marked by covering the aircraft with the label. When manual control started, I felt relieved that I could immediately start solving the future conflicts/problems.

F.4.2. Questions Answered Only by Participants in the Fictional Aircraft Group

What is your opinion on the number of fictional aircraft present on screen? Was it too little (1), too much (7) or balanced(4)?

Participant No.	**1**	**3**	**5**	**7**	**9**	**11**	**13**	**15**
Answer (1-7)	2	2	2	4	4	4	2	2

Please explain and elaborate on your previous answer.

- **Participant 1**
More fictional aircraft would keep me more awake.
- **Participant 3**
I think the scenario was too simple with the amount of fictional aircraft. It only took a few actions to direct them correctly and they almost never got into a nearly los situation
- **Participant 5**
In my opinion there should have been a few more fictional aircraft on screen so the proportion of time that you are solving conflicts instead of observing the sector increases. I feel like this would have increased my vigilance considerably.
- **Participant 7**
I think it was exactly right. It allowed for close inspection of the rest of the real aircraft while maintaining engagement with the virtual ones.
- **Participant 9**
If they are going to be there anyway, this is a good balance, not yet keeping attention away from real aircraft
- **Participant 11**
I would say that the amount of fictional aircraft was just about right to keep the situation manageable, especially when taking into account the fact that checking the automation decisions for the real aircraft takes up quite a bit of time.
- **Participant 13**
there were very little to keep me engaged during the phase in which the automation was most boring, but at the same time i didn’t appreciate their presence because I knew controlling them was a secondary objective, they appeared to me as a distraction whenever a real situation would require my attention.
- **Participant 15**
I think more aircraft would have made me more vigilant because now there were very few cases in which the virtual AC interacted with the real ones, making it so that even though they were there, I had to mind them very little.

How did you find the distribution of fictional aircraft in the sector to be? Were they well (7) or badly distributed (1)?

Participant No.	1	3	5	7	9	11	13	15
Answer (1-7)	6	5	6	4	6	6	4	6

Please explain and elaborate on your previous answer.

- **Participant 1**
They were coming from various directions, so nicely distributed. Maybe the only thin that could make it even more if there would've been two fake aircraft close to each other.
- **Participant 3**
They seemed to be randomly distributed, which is nice to keep your attention divided over the whole screen
- **Participant 5**
The distribution of fictional aircraft was OK for me. They spawned from multiple directions and followed roughly the same distribution as real aircraft in my opinion.
- **Participant 7**
They allowed me to train on getting the aircraft through many different routs. This definitely made it easier after the automation failure because I could plan and predict ahead better.
- **Participant 9**
It wasn't always the same aircraft with the same track
- **Participant 11**
A good mix of different fictional aircraft going towards different exit waypoints, with some pre-programmed conflicts as well.
- **Participant 13**
I think they were uniformly distributed, but i didn't feel their distributions as an important factor
- **Participant 15**
I think they were nicely distributed in a versatile way, not much to add.

How did you see and treat fictional aircraft? Were they more of a tool/game element or did you see them the same as actual traffic? Please elaborate.

- **Participant 1**
I mostly saw them the same. Actually even as more important to watch because I had to make sure real aircraft don't crash into them.
- **Participant 3**
I treated them as actual aircraft, I guess that that happens automatically when they also behave and have the same restrictions as the actual aircraft. I wonder if in a real life situation people would give them less priority

•Participant 5
I didn't really see them as purely a game element. I tried to treat them like actual traffic and act accordingly. I did however tried to give solving these conflicts lower priority than spotting anomalies with real aircraft, which sometimes proved to be more difficult than anticipated...

•Participant 7
I tried to see them as real aircraft, but I was willing to experiment more with them. This experience made some close calls easier to solve after automation failure.

•Participant 9
mostly as real traffic, unless this would come at the cost of actual traffic.

•Participant 11
I saw the fictional aircraft more as a game element, because I knew that the were not there in real life.

•Participant 13
game element. I didn't see them as traffic but just as a secondary element, which kept me engaged in something which wasn't the control of the automation, a disengagement on the actual task, a distraction.

•Participant 15
I saw them as actual traffic, really making sure they wouldn't lose separation. However I did treat them differently in a sense that I felt a lot better about changing their heading multiple times.

Do you feel like fictional aircraft helped you in remaining vigilant and engaged? Please elaborate.

•Participant 1
Yes! Actually controlling aircraft makes a lot of difference. It also gives you a good idea of the response times of aircraft etc. so it didn't feel like a big step when I had to take over.

•Participant 3
A bit, it helped me to concentrate on the whole simulation, it is good to have different tasks than just supervise and report

•Participant 5
Certainly! I feel like without the virtual aircraft it would have been more difficult to stay vigilant since you would be purely acting as an observer all the time until something actually goes wrong. It is also great to practice your conflict solving skills for situations were the automation fails and you have to suddenly control all aircraft. I feel like it sometimes did take some of my attention away from checking if any anomalies were happening with real aircraft though...

•Participant 7
I dont think in the time frame of the experiment it made much of a difference. But I do feel that they had a crucial function in keeping the skills up to date. I also feel strongly that for longer periods of time, they would certainly help.

•Participant 9
Not really, although this might be different when you do this day in day out.

•Participant 11
I am not sure. I guess it did help me to be more vigilant because you are actively involved in the loop of some aircraft. I do not know if I was engaged more because of the fictional aircraft, because they do not help me in creating a mental model of the traffic situation with only the real aircraft.

•Participant 13
I had the feeling of being engaged in a different task compared to the one i had at hand.

•Participant 15
I think they helped a little bit yes, but as mentioned in previous answer, I think that there were too few to actually have to keep them in mind all the time.

What is your opinion on the overall implementation of fictional aircraft?

•Participant 1
Promising, but would require some more thoughts into how many fake aircraft you need. Also, it does feel a bit 'useless' to control fake aircraft when you could just as well control the real aircraft :-)

•Participant 3
It might be useful. I guess there is a perfect point between helping to concentrate and being too distracted by too many fictional aircraft

•Participant 5
I found the implementation to be quite helpful. As indicated earlier, I feel like some slightly more complex situations to resolve would have helped staying vigilant slightly. The fictional aircraft were depicted in a way which made it trivial to distinguish them from real aircraft, which is a must in my opinion.

•Participant 7
It was near perfect for me.

•Participant 9
I did not have the impression that they helped me stay alert, but otherwise the balance of number of aircraft and added complexity was good.

•Participant 11
I don't really like the presence of fictional aircraft. When working with them, you have to accept the fact that not cooperative solutions between the fictional and real aircraft are possible (and there the automation controlling the real aircraft can sometimes work 'against you'). Furthermore, the fictional aircraft disappear as soon as the automation fails, slightly distorting the mental model of the traffic situation you have built up. This can (in some situations) make it quite difficult to adapt to the new situation without automation.

•Participant 13
I think their presence wouldn't work to keep people engaged in the supervision of the automation task, and just provide a distraction. I would study another method to keep people engaged.

•Participant 15
I think it was pretty nice, although perhaps in training it would be nice to have a situation where the virtual aircraft would lose separation with a real aircraft, to really drive home that it doesn't matter if they do. I think it's pretty easy to forget that they are just a game.