Gamification for Increased Vigilance and Skill Retention in Highly Automated Air Traffic Control

Călin Andrei Badea, Gijs de Rooij, Clark Borst, Max Mulder
Control & Simulation, Aerospace Engineering, Delft University of Technology, The Netherlands

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

I. 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 (ATC) domain is predicted to undergo fundamental modernisation in the next 20 years, as the push for increased automation is gaining traction. Controller support tools are currently being used, such as trajectory prediction and short-term conflict alerting [1], however the decision-making process still rests upon the air traffic controller (ATCO).

The SESAR Air Traffic Management Master Plan of 2020 [2] mentions that a high level of automation in ATC should 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 (CD&R) systems implies that ATCOs will undertake the role of supervisors, intervening only at the occurrence of exceptional events, such as automation failures. This will have major implications on the work environment, as the nature of the ATCO's task fundamentally changes.

The Ironies of Automation [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 Parasuraman 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. 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 improves situation awareness, even if conflict resolution is fully automated. Borst et al. [7] propose the use of ecological interfaces to help system supervisors detect faults in automation by increasing transparency and information flow within the human-machine system. Pop et al. [8] showed that producing and maintaining engagement benefits failure detection in ATC supervisors, and is a key element towards improved cognitive abilities.

However, one issue remains under-addressed: in a highly reliable and highly-automated system, automation failures will be rare, and maintaining engagement in such an environment is more challenging when the intervention of operators is rarely required. Munyung et al. [9] showed that keeping ATCOs engaged in a highly-automated environment by means of asking task-related questions at regular time intervals can lower workload peaks experienced in case of automation failure, but that did not lead to improved manual control performance once automation failed. One potential way to keep controllers both cognitively engaged and skilled could be through the use of gamification (the use of game elements in non-gaming contexts [10]). Threat image projection (TIP), 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, mitigating the effects of boredom [11].

This paper proposes the adaptation and implementation of TIP within a highly-automated en-route ATC environment through the use of fictional flights superimposed on real traffic, the latter being controlled by a fully automatic CD&R system. Fictional flights would require manual control (mitigating skill erosion) and the development and maintenance of a mental model of the flights in the sector (maintain situation

sesar JOINT UNDERTAKING 12th SESAR Innovation Days 5-8 December 2022, Budapest HungaroControl sesar DIGITAL ACADEMY

awareness), thus inciting higher engagement while mitigating the effect of boredom. The increase in engagement is expected to have a positive effect on supervisory control performance, and seeks to improve the transition from supervisory to manual control in case of automation failure by maintaining the operator in the control loop. This type of tool could be beneficial in low-workload situations (such as night time), helping maintain vigilance and focus within the work environment.

II. Design and Implementation

The main feature of TIP is that fictional threats are introduced to the work space of a system supervisor. Within the ATC domain, the most prominent threat to the overall safety level of operations are aircraft conflicts. The proposed implementation of gamification revolves around the use of fictional flights 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 flights, which would be controlled by automation, fictional flights must be manually commanded by system supervisors.

Fictional flights were implemented such that operators would be aware of their nature and that they were distinguishable from real flights. This increased transparency compared to TIP, where operators are not initially aware of the fictional nature of threats. It also facilitates the differential prioritisation of tasks. Both supervising real flights and manually controlling fictional flights share the same goal, routing flights through the sector safely and efficiently. By considering the display design principles of discriminability and redundancy described by Wickens [12], the design presented in Figure 1 was created, in which fictional flights differ from real flights in both colour and shape. The colour blue was chosen due to its high contrast with the dark backgrounds often used in ATC displays.

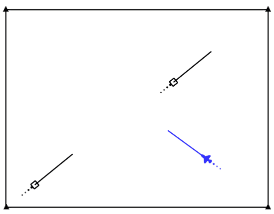


Figure 1. Visualisation of a fictional flight introduced on the ATC radar screen.

By analysing the design considerations of threat image projection [13], two main factors that affect the influence of fictional flights on operators can be determined: the complexity of the traffic situation and the number of threats present on the screen. The first can be correlated with the nature of the conflicts induced by fictional flights (e.g., head-on, catching up) which can be influenced by their traffic pattern. If an airspace sector is assumed to contain several main routes along which most flights will travel, there are three types of traffic patterns relative to these routes: (1) fictional flights are introduced such that they cross the main traffic flow, (2) fictional flights are introduced among the flights of the main flow, or (3) as a combination of among and crossing the main flow. The second influencing factor is correlated with the number of fictional flights present in the sector at one time, which can influence the workload and concentration that an operator must allocate towards them. A high number of fictional flights demands higher workload, thus making the supervisory task more difficult to perform. This is undesirable, as supervising air traffic is the highest priority.

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

III. Method and Experimental Setup

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

The experiment was designed to be performed in a controlled and simplified ATC 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 were given the task to report anomalous events that might occur when automation is in charge of controlling flights. Participants would also have to intervene when automation would experience a failure, thus transitioning from supervisory to manual control of flights. Thus, the experiment scenario run was divided into two phases: a supervisory control phase and, after the failure event, a manual control phase.

A. Apparatus and software

The experiment was conducted using SectorX, a TU Delft in-house developed Java-based ATC simulator, that was modified to include fictional flights, a supervisory control mode and a manual control mode. It was run using the hardware setup presented in Figure 2. 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. Separation circles with a radius of 2.5 nautical miles, history dots, and one minute look-ahead

Figure 2. The hardware setup used for the experiment.

velocity vectors were added to the aircraft blips to facilitate the supervision task.

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

The simulator was augmented to accommodate anomaly reporting during the supervisory control phase. While automation was active, unusual behaviour of real flights could be reported by clicking the offending flight and typing a report in a pop-up window. 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 flights was enabled.

B. Participants

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

C. Participant tasks

Participants were instructed to perform the following tasks:

TABLE I. PARTICIPANT CHARACTERISTICS.

	Fictional flights	Control
Expertise		
Master students	5	4
Doctorate students	2	3
Lecturers	1	1
	8	8

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

D. Air traffic scenarios

A clipped version of the Delta sector of Maastricht Upper Area Control Centre (MUAC) was selected for the experiment due to the familiarity of the participants with the Dutch airspace, data availability for this sector, and because it is an en-route sector. ADS-B data from the year 2018 were analysed and used to develop realistic traffic patterns.

Three restricted airspaces were placed at the edges of the sector to test the situation awareness of participants with events occurring away from the centre of the screen. Entry and exit waypoints were distributed along the sector boundary, as well as three inner waypoints coinciding with the high traffic areas. The final sector configuration is presented in Figure 3.

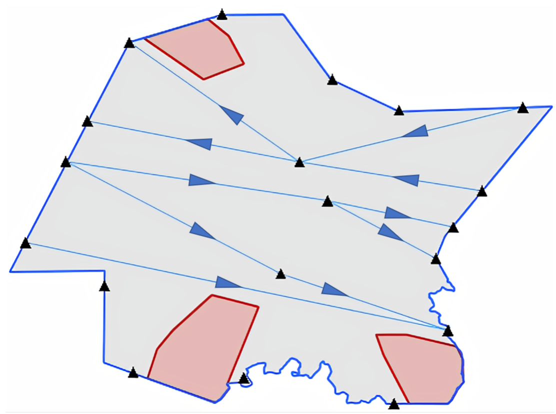


Figure 3. Modified Delta sector used for the experiment traffic scenario.

The baseline scenario had a traffic density of 45 flights per hour, with approximately eleven aircraft within the sector at all times. The length of the scenario was based on the attention decrement phenomenon described by Mackworth [14], who experimentally showed that the greatest decrement in attention, while performing a monotonous supervisory task, occurs in the first hour. Hancock [15] expands on the concept and links the attention decrement to the nature of the task. Based on these considerations, the length of the supervisory control part of the scenario was set at 4,000 seconds (66.7 minutes) and the length of the manual control part at 700 seconds (11.7 minutes).

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

E. Automation, anomalies and failure

Traffic scenarios included pre-programmed commands such that real flights would maintain a separation of approximately 7 nautical miles from each other, follow the routes presented in Figure 3, and manoeuvre at the three interior waypoints when possible to increase automation transparency and predictability. No other automated or aiding tool (such as conflict detection) was enabled during the experiment. As the commands were pre-programmed, the presence of fictional flights had no impact on the commands given by automation, thus conflicts could arise between fictional and real flights that needed human intervention.

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

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

In total, seven anomalous events occurred at the time intervals presented in Table II. Two anomalies occurred 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 drawn towards one part of the screen).

TABLE II. ANOMALOUS EVENTS WITHIN THE EXPERIMENT SCENARIOS.

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

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

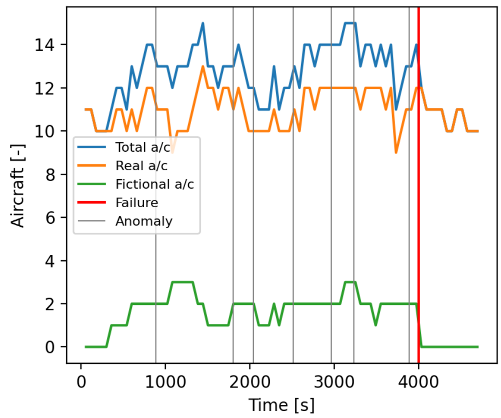


Figure 4. Summary of fictional flights and baseline scenarios.

F. Experiment procedure

The procedure used for each individual experiment session is presented in Figure 5. After undergoing a short briefing session, participants underwent a series of training scenarios both with and without fictional flights before being informed of their group assignment. At the end of the supervisory phase, after an automation failure, the participants needed to manually control flights. The experiment was concluded with a survey.

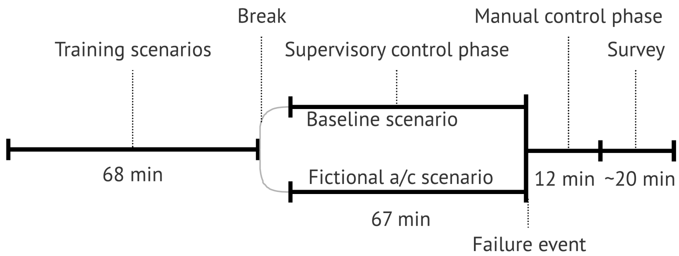


Figure 5. Experiment timeline. All participants undertook the same training, and the same setup for the manual control phase. The supervisory control phase differed depending on experimental group.

G. Control variables

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

- *Real traffic:* All participants experienced the same real traffic during both the supervisory and manual control phases of the scenarios, including the automation commands.
- *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 with constant indicated airspeeds between 250 and 310 kts.
- *Anomalous events:* Automation anomalies were the same across the fictional flights and baseline scenarios.

- *Radar update rate:* The radar update frequency was set at once every 5 seconds (0.2 Hz).

H. Dependent measures

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

$$\text{CRR} = \frac{\text{Number of clicks after failure}/700[s]}{\text{Number of clicks before failure}/4,000[s]} \quad (1)$$

- **Subjective questionnaire**
  - *Situation awareness:* using SART questions [16].
  - Several Likert-scale and open questions about control strategies, order of priorities when supervising automation, experience with and trust in automation and experience with fictional flights, if present.

I. Hypotheses

The hypotheses of the experiment were the following:

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

HP-2 The use of fictional flights within a highly automated ATC environment improves the vigilance levels of operators.

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

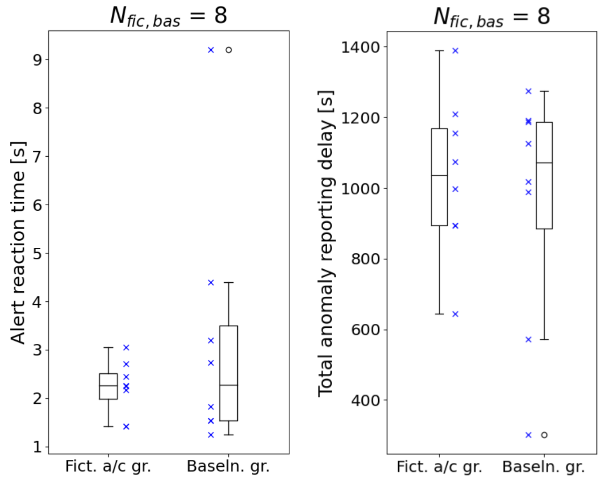
HP-4 The presence of fictional aircraft within an ATC supervisory control environment improves situation awareness of operators.

IV. Results

This section first presents the objective measures in terms of supervisory and manual control performance, engagement and subjective questionnaire results. Due to the small sample size (eight per group), Mann-Whitney U tests were used to analyse the data.

A. Supervisory control performance

Starting with supervisory performance, Figure 6a portrays the reaction times of participants to the automation failure notification. While the medians for the two groups are identical, the baseline group's reaction time shows a larger spread.



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

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

Figure 6. Supervisory performance per experiment group.

The outlier within the baseline group is a participant who inadvertently attempted to switch to manual control without first dismissing the alert. No significant difference between the two groups is observed.

The same remark can be made when considering the total anomaly reporting delay per participant during the supervisory phase, presented in Figure 6b. The variance of the dataset is relatively large due to the diversity in participant strategy when reporting anomalies, which they were told to only report when being confident about them.

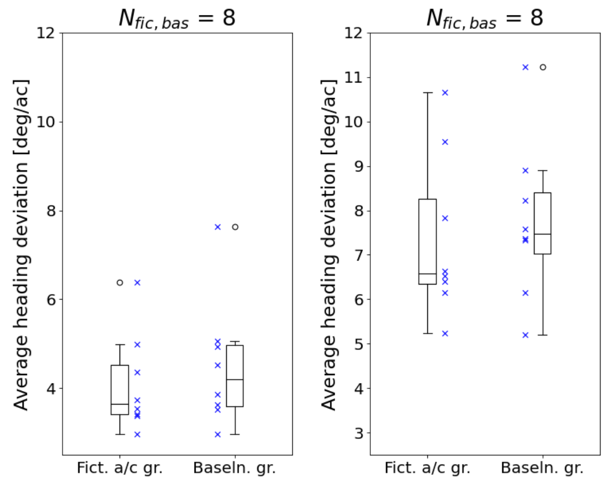
It should be mentioned that all participants detected all anomalies that occurred during the scenario. However, a notable result can be observed when considering false positive anomaly reports (mostly consisting in flights manoeuvring late): most of these were submitted by participants in the fictional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

B. Manual control performance

Moving on to manual control performance, the average heading deviation does not reveal a significant difference either, although the median of the fictional flights group is higher both over the whole time interval after automation failure (Figure 7a) and within 2 minutes after failure (Figure 7b). A relatively high variability in the data set can also be observed, as participants used different strategies when manually controlling flights. It should be noted that the heading coefficient does not capture all aspects of performance, and in essence represents the time efficiency with which participants solved the immediate conflicts after failure.

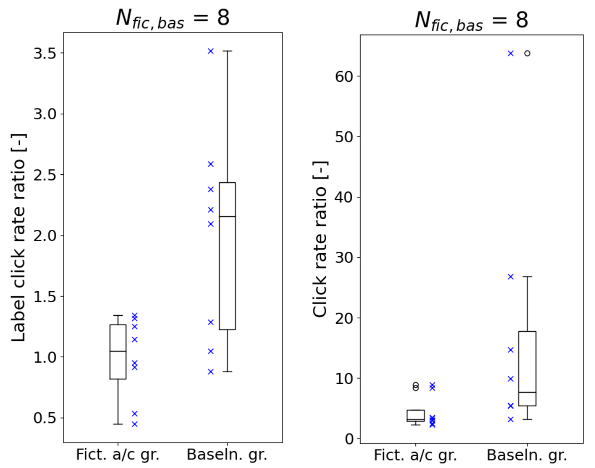
C. Engagement

In terms of engagement, a significant difference between the experimental groups was observed in the ratio between the label click rate before and after automation failure ($N = 8$, $U = 12$, $p = .038$). From Figure 8a, it can be seen that



(a) Throughout manual control (b) First 120 seconds after failure

Figure 7. Manual control performance (average heading deviation) per group.



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

Figure 8. Mouse click rate ratios split per group.

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

The difference in click rates between the two groups can also be seen when the cumulative number of mouse clicks is plotted over time, as presented in Figure 9. Due to technical difficulties, this timestamped click data was only recorded for five participants in each group. The activity of participants seems to be similar in the first 1,000 seconds of the simulation

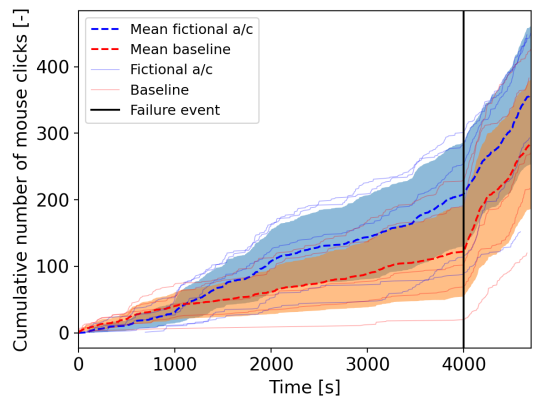


Figure 9. Cumulative mouse clicks over time for five participants per group. Also shown are the averages per group and their 95% confidence intervals.

when fictional flights are not initially present on screen. Then the activity of the fictional flights group is higher after this mark as interactions between fictional and real flights become more apparent. Furthermore, there is an initial peak of fictional flights between 1,000 and 2,000 seconds, as seen in Figure 4. However, shortly after the failure event (between 4,000 and 4,200 seconds), the average click rate of the baseline group is higher, suggesting that the change in mouse activity is more sudden among this group than the fictional flights group.

D. *Subjective questionnaire data*

This section presents the subjective data obtained from the survey at the end of the experiment, focusing on the questions that participants had to answer using a 1-7 Likert scale. The SART index computed from the answers did not yield a significant difference between the fictional flights and the baseline groups. However, among the individual SART questions, differences can be observed in the concentration level and information usefulness ratings, presented in Figure 10.

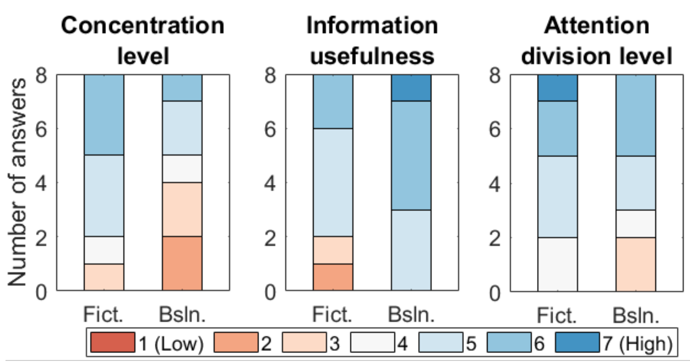


Figure 10. Answers to selected SART questions represented on a colour-coded Likert scale.

The answers for information usefulness also reveal a trend: participants in the baseline group indicated that they had a greater understanding of the knowledge received from the display, whereas the fictional flights group scored lower overall. A difference between the two groups can also be observed in the answers for the attention division level question: participants

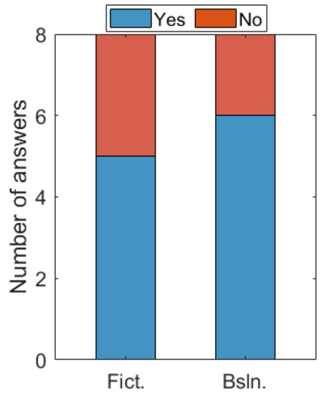


Figure 11. Answers to the question "*Do you think the presence of fictional flights helped/would have helped you to be more vigilant?*".

in the fictional flights group overall reported higher levels, which is expected due to there being more elements on the screen that require attention.

Finally, each participant was asked whether they found or would have found fictional aircraft beneficial for maintaining vigilance within the experiment environment. Overall, more than half of the participants in both groups considered that fictional flights were or would have been useful (Figure 11).

V. Discussion

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

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

The threshold required for a significant attention and performance decrement was not attained, as all participants reported all intended anomalies presented in Table II 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. This is confirmed within the open questions answered by participants after the experiment: most participants from both groups reported that the occurrence of anomalies, as well as the beforehand knowledge of their existence, contributed positively towards maintaining vigilance, which enabled better anomaly detection performance. Thus, no conclusion can be drawn relating to hypothesis **HP-1** (fictional flights use improves anomaly detection rates) and **HP-2** (fictional flights use improves vigilance and reaction times), as no detection misses occurred. Moreover, the manual control performance measures (e.g., Figures 7a and 7b) show a high degree of variability with no significant difference between groups. Thus, conclusions cannot be drawn regarding hypothesis **HP-3** (fictional flights use improves manual control performance after automation failure).

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

The false-positive reports submitted by some participants provide some insight into the effects of the presence fictional aircraft. While not being considered anomalies, these events require the attention and supervision by participants during the scenario. Thus, false-positive reports could be an effect of an overall increase in engagement and situation awareness achieved by the presence of fictional flights, resulting in the detection of borderline anomalous events that were mostly not reported by the baseline group. This comes in support of hypothesis **HP-4** (the fictional flights group participants achieved higher levels of situation awareness).

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

Finally, the results from the SART questionnaire show that participants in the fictional aircraft group experienced a better concentration level (associated with mental workload), thus supporting hypothesis **HP-4** (the presence of fictional flights 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. However, the trend observed in the information usefulness answers shows one of the dangers in overlaying fictional flights on top of real flights: the screen itself contains more information, however this information does not contribute to understanding the (real) situation. This is backed by some participants in the fictional flights group reporting that they found fictional flights distracting.

VI. Conclusion and Recommendations

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

The subjective results obtained in this paper show that the use of fictional flights 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, a tool that uses this strategy to enhance cognitive abilities could be implemented such that it can be toggled on and off. This would allow controllers to enable and disable the tool depending on the real traffic situation and personal preference. Such a tool could also be implemented using objective measures, as proposed by Di Flumeri et al. [17], where the number of fictional aircraft can be adjusted in function of the measured vigilance level.

Further research should be performed to better understand the potential benefits of using such tools and discover ways through which the negative effects could be mitigated. Professional ATCOs should be used for better understanding of how cognitive skills are influenced by the proposed tool in a more realistic ATC environment. Furthermore, future experiments should collect more data (e.g., eye tracking) to improve measurements quantifying attention and vigilance. Other game elements and strategies should also be explored (e.g., scores and achievements) 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. Thus, investigations should be performed on whether the same vigilance improvement effects can be obtained in ways more compatible with current ATC work environments. De Rooij et al. [18] proposes automating only part of air traffic, with part of them still requiring conventional manual control. This might have the same effect as fictional aircraft, but without introducing new elements to the work environment.

References

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

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

[3] L. Bainbridge. "Ironies of automation". In: *Automatica* 19.6 (1983), pp. 775–779. 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 (2018). 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). 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: *13th IFAC Symposium on Analysis, Design, and Evaluation of Human-Machine Systems* (2016). DOI: 10.1016/j.ifacol.2016.10.458.

[7] 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 (2015). DOI: 10.1109/THMS.2014.2364984.

[8] 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.

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

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

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

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

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

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

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

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

[17] G. Di Flumeri, F. De Crescenzio, B. Berberian, O. Ohneiser, J. Kramer, P. Aricò, G. Borghini, F. Babiloni, S. Bagassi, and S. Piastra. "Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems". In: *Frontiers in Human Neuroscience* 13 (2019). DOI: 10.3389/fnhum.2019.00296.

[18] G. de Rooij, C. Borst, M. M. van Paassen, and M. Mulder. "Flight Allocation in Shared Human-Automation En-Route Air Traffic Control". In: *International Symposium on Aviation Psychology*. 2021, pp. 172–177.