Human-Automation Flight Allocation in Shared Air Traffic Control

de Rooij, G.

DOI
10.4233/uuid:15496a40-3329-4ed5-bf8f-d366bafe5ded

Publication date
2025

Document Version
Final published version

Citation (APA)
de Rooij, G. (2025). *Human-Automation Flight Allocation in Shared Air Traffic Control*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:15496a40-3329-4ed5-bf8f-d366bafe5ded

HUMAN - AUTOMATION FLIGHT ALLOCATION IN SHARED AIR TRAFFIC CONTROL

Gijs de Rooij

Human-Automation Flight Allocation in Shared Air Traffic Control

Human-Automation Flight Allocation in Shared Air Traffic Control

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Wednesday 21 May 2025 at 15:00 o'clock

by

Gijs DE ROOIJ

Master of Science in Aerospace Engineering
Delft University of Technology, Delft, the Netherlands
born in Amsterdam, the Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. ir. M. Mulder,	Delft University of Technology, promotor
Dr. ir. C. Borst,	Delft University of Technology, promotor
Independent members:	
Prof. dr. ir. J.M. Hoekstra	Delft University of Technology
Prof. dr. S.C. Pont	Delft University of Technology
Prof. dr. M.S. Young	University of Southampton, United Kingdom
Dr. J. Blundell	Cranfield University, United Kingdom
Dr.-Ing. B.R. Korn	German Aerospace Center, Germany

Dr. ir. M.M. van Paassen and A.B. Tisza have contributed greatly to the realization of this dissertation.

TU Delft Delft University of Technology

Keywords: air traffic control, automation, human-autonomy teaming

Printed by: Ipskamp Printing

ISBN 978-94-6384-788-9

An electronic version of this dissertation is available at repository.tudelft.nl.

If you build systems where people are rarely required to respond,
they will rarely respond when required.

Hancock (2014, p. 453)

Summary

Air traffic control (ATC) is transitioning towards a more automated system where human air traffic control officers (ATCOs) are increasingly supported by systems working at a high(er) level of automation (LOA). Made possible by advancements in computing power, artificial intelligence and a more data-driven air traffic management (ATM) system, automation is expected to address major issues, such as a global staff shortage, growing air traffic demand and environmental concerns.

On this shift towards greater reliance on automation, two main strategies can be identified that each have a distinct impact on the system's operators (i.e., ATCOs). **Chapter 2** details how these differ between a traditional function-based strategy, where all flights are controlled at a gradually increasing LOA, and a constraint-based strategy, where a subset of flights is operated at a higher LOA than other flights. The former strategy brings many human-automation issues that have been widely demonstrated through empirical research, such as 'out-of-the-loop' situation awareness, transient workload peaks, skill erosion, boredom and reduced job satisfaction. The latter strategy has the advantage of avoiding mixed authority over individual flights by creating a more parallel system than the function-based serial system. The resulting human-autonomy team (HAT) accelerates the introduction of higher LOA in operational environments, fostering innovation.

The HAT perspective has only recently appeared on the radar of the ATC community, and practical examples of its potential and implications are scarce. An interesting example is found at Maastricht Upper Area Control Centre (MUAC), an air navigation service provider (ANSP) responsible for air traffic above 24,500 ft over Belgium, Luxembourg, the Netherlands, and part of Germany. MUAC is currently employing a constraint-based strategy in the development of a future shared airspace where ATC services for low-complexity routine flights are fully automated while complex flights stay with the ATCO. A key challenge for such an ATC system is to determine which flights should be allocated to either the human ATCO or the automation.

This research set out to broaden the knowledge about constraint-based automation in ATC and the desired allocation of flights in particular. Each chapter addresses a sub-question, often through empirical research with professional MUAC ATCOs. The research had three phases, starting with a first exploration, followed by an impact analysis of flight allocation on ATCO workflows and the role of flight complexity in this. The thesis concludes with a validation exercise consolidating all insights from the preceding chapters.

To test several preconditions and general ATCO acceptance of this novel concept, **Chapter 3** begins with an exploratory simulator experiment. The participating ATCOs had full control over which flights they would delegate to the automation. Although predefined suggestions were presented, the ATCOs mostly ignored these. This experiment demonstrated the potential for allocating selected flights to either human or automation in a single airspace, but also stressed the importance of using a clever algorithm to determine this allocation. Geographic sector-based allocation, with automation handling all traffic in one sector and the ATCO all traffic in another sector, was rejected by the majority of participating ATCOs. They preferred an interaction-based allocation, hinting at the need to establish a complexity-score for each single flight.

Diving deeper into the impact that flight allocation might have on the workflow of an ATCO, **Chapter 4** focuses on the core ATCO tasks: conflict detection and resolution (CD&R). Following a literature study and on-the-job ATCO observations, cognition flowcharts were constructed for these two tasks. Through an experiment with simplified static traffic scenarios, in which ATCOs had to detect and resolve conflicts, the most cognitively demanding types of traffic situations were searched for, as a means to quantify the various cognitive paths that can be traversed in the flowcharts. This turned out to be challenging, as ATCOs, like other experts, make frequent use of shortcuts and parallel processing. The constructed flowcharts can, however, serve as a starting point for the design of more human-like CD&R algorithms, such as used in this thesis' experiments. Automation that performs tasks in similar fashion as an ATCO might increase operator acceptance. This chapter's results stressed the importance of understanding flight-centric complexity before the impact of flight allocation on workflows can be determined.

To increase this understanding, the experiment in **Chapter 5** used actual traffic snapshots overlaid with a single flight of interest for which the ATCOs had to indicate their perceived complexity. This individual flight complexity was a unique approach, compared to existing literature that mainly considers sector-wide complexity. Despite individual differences, flights on either end of the complexity scale were reliably identified. These results indicate that a flight allocation scheme may not need to be fine-tuned towards individual ATCO preferences. In general, a flight's complexity appears to be mostly driven by (potential) spatiotemporal interactions with other flights.

Consolidating the insights from preceding chapters, **Chapter 6** discusses the most realistic and extensive experiment of this thesis. It replicates the experiment from Chapter 3 while addressing many of that experiment's shortcomings. Lessons learned in the preceding chapters led to several improvements, such as an increase in automation capabilities and communication, and more informed allocation schemes than the pragmatic schemes from the first experiment. In a direct comparison between two distinct allocation schemes, it was found that an interaction-based scheme is subjectively preferred by ATCOs and shows small efficiency benefits over a simpler flow-based allocation. In addition, it was concluded that automation should be sufficiently equipped to issue the same instructions as ATCOs, and should have the same notion of constraints from letters of agreement, to create a common ground and reduce mixed conflicts.

In conclusion, this thesis has brought forward the knowledge about flight allocation in an airspace that is shared between a human ATCO and a computer system. It can serve as a starting point for future research and development of highly automated ATC systems. Fully autonomous ATC will not become a reality in the short-term, but results show promising effects and a general feasibility of higher LOA applied to a constrained environment (i.e., a subset of flights). Researchers and ANSPs are encouraged to step beyond purely function-based visions on automation allocation and embrace a constraint-based automation strategy. This thesis has shown that a combination of these two strategies may lead to desired human-automation teamwork.

Samenvatting

Luchtverkeersleiding (ATC) evolueert naar een steeds hoger niveau van automatisering (LOA) waarin luchtverkeersleiders (ATCOs) door steeds meer systemen ondersteund worden. Dit wordt mogelijk gemaakt door vooruitgang op het gebied van rekenkracht, kunstmatige intelligentie en een meer datagedreven ATC-systeem. Automatisering wordt gezien als (deel)oplossing voor enkele grote wereldwijde problemen, zoals personeelstekort, de groei van het luchtverkeer en het klimaatprobleem.

Op weg naar hogere LOAs kan men twee strategieën onderscheiden, zoals beschreven in **Hoofdstuk 2**. Bij een traditionele functiegebaseerde strategie worden alle vluchten geleidelijk op een steeds hoger LOA afgehandeld. Dit leidt tot allerhande problemen tussen mens en automatisering die veelvuldig zijn aangetoond in empirisch onderzoek, zoals een verminderd toestandsbewustzijn ('out-of-the-loop'), kortstondige werklastpieken, verlies van vaardigheden, verveling en verminderd werkplezier. Bij een voorwaardengebaseerde strategie wordt slechts een deel van de vluchten op een (nog) hoger LOA afgehandeld. Dit heeft als voordeel dat er sprake is van een meer parallel systeem, waarin de verantwoordelijkheid over afzonderlijke vluchten bij de mens danwel de automatisering ligt. Het resulterende mens-automatiseringsteam (HAT) versnelt de introductie van hogere LOAs in operationele werkomgevingen en bevordert zo innovatie.

Het HAT-perspectief is pas recent op de radar van de ATC-gemeenschap verschenen waardoor goede praktijkvoorbeelden schaars zijn. Een interessant voorbeeld is te vinden bij Maastricht Upper Area Control Centre (MUAC), een instantie die verantwoordelijk is voor het luchtverkeer dat op 24,500 voet of hoger boven België, Luxemburg, Nederland en een deel van Duitsland vliegt. Op basis van een voorwaardengebaseerde strategie ontwikkelt MUAC een gedeeld luchtruim waar simpele routinevluchten in de toekomst door een computer afgehandeld worden, terwijl complexe vluchten bij de ATCOs blijven. Eén van de belangrijkste uitdagingen van zo'n voorwaardengebaseerd systeem is om te bepalen welke vluchten aan de mens of aan de computer moeten worden toegewezen.

Dit promotieonderzoek had als doel om de kennis over voorwaardengebaseerde automatisering in een ATC-context te vergroten, waarbij de focus op het toewijzen van vluchten ligt. Elk hoofdstuk behandelt een deelvraag, vaak door middel van empirisch onderzoek met professionele ATCOs van MUAC. Het onderzoek bestond uit drie fasen en begon met een eerste verkenning, gevolgd door een analyse van veranderende ATCO-taken bij het invoeren van individuele vluchttoewijzing en de rol van vluchtcomplexiteit hierin. Het proefschrift eindigt met een validatie-experiment waarin de lessen uit voorgaande hoofdstukken samenkomen.

Om de vooroordelen en acceptatie van ATCOs omtrent dit nieuwe concept te peilen werd een eerste simulatie-experiment uitgevoerd (**Hoofdstuk 3**). De deelnemende ATCOs hadden de volledige controle over welke vluchten ze aan de automatisering toewezen. Hoewel er suggesties werden getoond negeerden de meeste ATCOs deze. Dit

experiment demonstreerde de potentie van het gedeelde-luchtruimconcept, maar benadrukte ook het belang van een slim toewijzingsalgoritme. Toewijzing op basis van geografische sectoren, waarbij de automatisering alle vluchten in een bepaald gebied afhandelt en de ATCO alle vluchten in een ander gebied, werd door de meeste ATCOs afgewezen. Zij prefereerden toewijzing op basis van onderlinge interacties tussen vluchten, wat wijst op de noodzaak om de complexiteit van individuele vluchten te bepalen.

Om de impact van vluchttoewijzing op het werk van ATCOs te specificeren, focust **Hoofdstuk 4** op twee hoofdtaken: conflicten detecteren én oplossen (CD&R). Een literatuurstudie en werkplekobservaties resulteerden in cognitieve stroomdiagrammen voor deze twee taken. Middels een experiment met versimpelde statische verkeerssituaties, waarop ATCOs CD&R moesten toepassen, werd gezocht naar de meest veeleisende situaties om zo de stromen in de diagrammen te kunnen kwantificeren. Dit bleek lastig te zijn omdat ATCOs, net als andere experts, regelmatig aan parallelle verwerking doen en stappen overslaan. De stroomdiagrammen kunnen echter als startpunt dienen voor meer mensachtige CD&R-algoritmes zoals geïmplementeerd in Hoofdstuk 3 en 6. Als een computer taken op eenzelfde manier als een ATCO uitvoert kan dit de acceptatie verhogen. De resultaten in dit hoofdstuk benadrukken het belang van complexiteitsbepaling per vlucht voordat de impact van vluchttoewijzing op ATCO-taken bepaald kan worden.

Om deze complexiteitsbepaling te verbeteren werd in het experiment uit **Hoofdstuk 5** telkens één vlucht aan echte radarbeelden toegevoegd, waarna de ATCOs de complexiteit van deze vlucht beoordeelden. Een unieke aanpak, aangezien de meeste literatuur alleen naar sectorbrede complexiteit kijkt. Los van individuele voorkeuren werden vluchten aan de uiteinden van de complexiteitsschaal betrouwbaar geïdentificeerd. Dit toont dat een toewijzingsschema niet op individuele ATCOs hoeft te worden toegespitst. De complexiteit lijkt bovenal afhankelijk van interacties tussen vluchten in ruimte en tijd.

In **Hoofdstuk 6** komen de verworven inzichten samen in het meest realistische en uitgebreide experiment van dit proefschrift. Het borduurt voort op het experiment van Hoofdstuk 3, maar lessen uit eerdere hoofdstukken leidden tot een aantal verbeteringen. Zo zijn de capaciteiten en communicatie van de computer uitgebreid, en werden beter onderbouwde toewijzingsschema's gebruikt dan de pragmatische schema's uit het eerste experiment. In een directe vergelijking tussen twee verschillende toewijzingsschema's bleek een interactiegebaseerd schema de voorkeur te hebben van de ATCOs en tot een iets betere efficiëntie te leiden ten opzichte van een simpel stromingsgebaseerd schema. Ook bleek dat de computer dezelfde instructies moet kunnen geven als ATCOs en hetzelfde begrip van procedures moet hebben, zodat er sprake is van een gemeenschappelijke basis en gedeelde conflicten vermeden worden.

Concluderend heeft dit promotieonderzoek de kennis vergroot over vluchttoewijzing aan een ATCO óf een computer in een gedeeld luchtruim. Het kan als startpunt dienen voor verder onderzoek naar de ontwikkeling van hoogautonome ATC-systemen. Volledig autonome ATC zal niet snel bereikt worden, maar de uitgevoerde experimenten toonden een positief effect en de haalbaarheid van hoge LOAs bij individuele vluchttoewijzing. Onderzoekers en ATC-instituten worden aangemoedigd om verder te kijken dan puur functiegebaseerde automatisering en een voorwaardengebaseerde strategie te omarmen. Dit proefschrift heeft aangetoond dat het combineren van beide strategieën tot de gewenste samenwerking tussen mens en automatisering kan leiden.

Contents

1

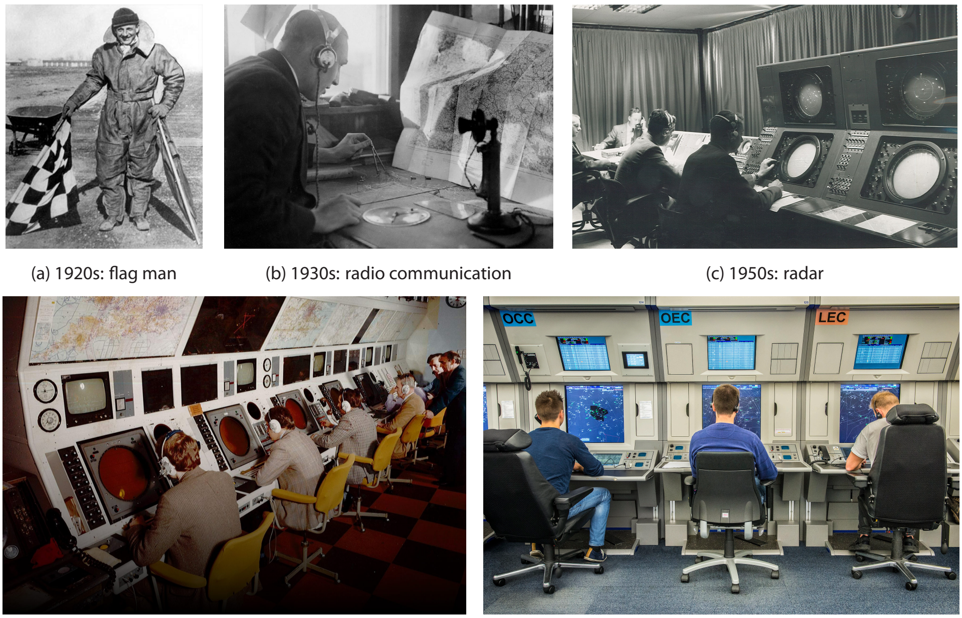
Introduction

This chapter introduces the current state of the art and future visions on automation in air traffic control. After listing some of the key challenges that generally come with an increase in automation, it continues with a discussion of several potential remedies that served as inspiration for this thesis. The chapter then presents the research goal and questions, as well as a definition of the scope and research assumptions. It concludes with an outline of the various chapters in this thesis and how they relate to each other.

Parts of this chapter have been published in:

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

In the early days of aviation, with just a handful of aircraft flying around only in daylight, there was no need to provide air traffic control (ATC) services. With increasing numbers of aircraft, airfields started using flagmen in the 1920s (Figure 1.1a) to signal pilots whether they could land or take-off (Nolan, 2011). The introduction of radio communication in the 1930s as shown in Figure 1.1b meant that aircraft could be controlled from farther away, and that more detailed instructions could be given (Gilbert, 1973). This increased even further after radar was invented in World War II, enabling the surveillance and control of aircraft far beyond the vicinity of the controller and in any weather or visibility condition (Figure 1.1c). In the following decades, area control centers (ACC) staffed by dozens of controllers, working in human-human teams, were constructed around the world to manage an ever-increasing stream of air traffic (Figures 1.1d and 1.1e).



(a) 1920s: flag man (b) 1930s: radio communication (c) 1950s: radar

(d) 1960s: en-route center (e) 2010s: modern en-route center

Figure 1.1: Evolution of air traffic control over the years (images by FAA, NATS and EUROCONTROL).

Looking back at the past century, one can see a substantial increase in the number of systems and humans involved. Nowadays, operators have a plethora of support tools at their disposal to help them handle the traffic, such as automated conflict detection and alerting. However, humans have not disappeared. In fact, the air traffic control officers (ATCO) remain at the very center of the system and play a key role in ensuring a safe, orderly and expeditious flow of traffic. Despite long-lived future visions on soon-to-be-operational highly automated ATC (Hunt and Zellweger, 1987) and abundant research in the area of automatic conflict resolution (Alaeddini et al., 2011, Erzberger et al., 2012, Frazzoli et al., 2001, Gariel and Feron, 2009, Trapsilawati et al., 2021), humans are still performing most of the work. Unlike in other parts of the aviation system, e.g., the flight deck of commercial airliners, automation is thus relatively sparse in ATC.

This is bound to change with the continuous quest for more efficient and safer air traffic management (ATM), driving the development of more advanced automation to support ATCOs. The Single European Sky ATM Research (SESAR) program and its United States equivalent Next Generation Air Transportation System (NextGen) both aim for higher levels of automation (LOA) in the coming decades, leading to a more supervisory and strategic role for humans (Joint Planning and Development Office, 2011, SESAR Joint Undertaking, 2020). In such an environment, fewer people are expected to handle more traffic in larger airspaces (Prevot et al., 2012).

1.1 Airspace sectorization

The number of flights in a (national) airspace is nowadays often so large that the airspace is divided into smaller sectors, with potentially an even further (dynamic) division in different altitude layers and sub sectors based on traffic demand (Baumgartner, 2007). This is today's most fine-grained flight allocation, shown in the top half of Figure 1.2.

An average commercial flight departing from an airport is first in contact with tower control, followed by approach and area control as it climbs higher and further away from the airport. Before entering the cruise phase, it may be transferred to upper area control. Every time the flight leaves a sector, the ATCO transfers the flight and asks the pilots to switch to the next sector's radio frequency. Depending on the flight's length and route, multiple en-route sectors are traversed, before starting the descent in which the flight is transferred between the aforementioned sectors in reverse order. En-route ATC, as executed by Maastricht Upper Area Control Centre (MUAC) above 24,500 ft over Belgium, Luxembourg, the Netherlands and northwestern Germany, is the scope of this thesis.

Each en-route sector is staffed by two ATCOs: an executive controller (EC) who communicates with pilots and has final authority over the sector, and a coordinating controller (CC) who coordinates with adjacent sectors to manage the EC's workload. Sector geometries are occasionally re-assessed and updated to reflect changing traffic demands and to (further) reduce the number of inter-sector conflicts. Research into dynamic sectorization aims to make sector geometries adapt in real time (Gerdes et al., 2018).

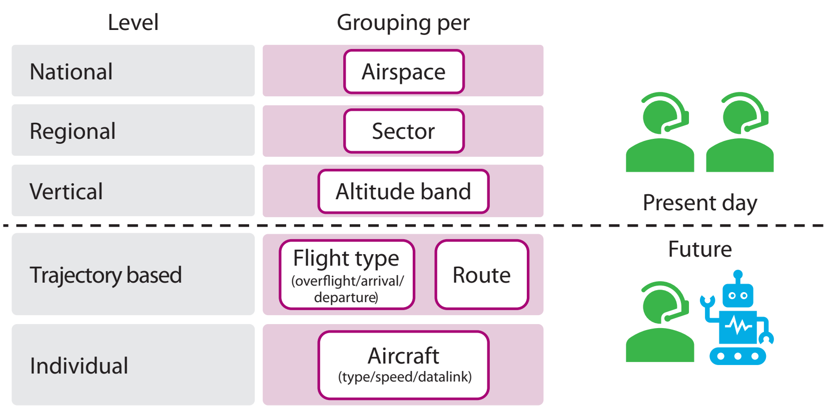


Figure 1.2: En-route control authority and responsibility allocation in current-day and future operations.

> *Note:* Throughout the figures and tables in this thesis, the color green is used when a reference is made to a human controller, while blue relates to the automated system. This is inspired by the colors on the current MUAC radar screen, where green flights are under control of the ATCO, whereas blue flights are within that ATCO's area of responsibility, but controlled by a different ATCO.

The planned introduction of trajectory-based operations (TBO), where flights strictly follow pre-negotiated trajectories, opens up the possibility of working with a more tailored division of responsibilities that is not tied to geographical areas, transitioning to the lower half of Figure 1.2. Potentially even down to the level of individual flights, such as proposed for flight-centric ATC (FCA, Volf, 2019), where one ATCO is responsible for the entire en-route trajectory of a flight. High bandwidth air-to-ground digital data links will enable this, greatly reducing the number of transfers between centers and balancing workload more evenly over the available ATCOs (Schmitt et al., 2011). The technology and capabilities that come with that also pave the way for the allocation of certain individual flights to an automated system, rather than automating an entire sector.

1.2 The rise of automation

The trend towards more (advanced) automation is not unique to ATC. In aviation, the increased use of automation on the flight deck is a prime example. In contrast to early aircraft designs with purely manual trajectory planning, guidance and control, modern airliners are largely automated. Pilots delegate control to the autopilot for prolonged flight phases and assume a largely supervisory monitoring role. Similar trends are seen in the automotive domain, where, after the adoption of driver assistance technologies like lane-keeping assistance, self-driving cars are on the horizon of several manufacturers (Merat et al., 2019). However, it is yet to be seen whether truly self-driving cars will (and should) ever be widely accepted.

Focusing on the European ATM system, Table 1.1 provides an overview of the various levels of automation (LOA) from the 2025 edition of the European ATM Master Plan (SESAR Joint Undertaking, 2024). Current-day operations are at Level 0, with the human ATCO responsible for most tasks, supported by automation in the information perception and analysis tasks (e.g., the display of routes and interpretation of raw radar data) and to some extent in the execution of actions. At Levels 1 and 2, the automation will assist the human in making decisions and selecting appropriate actions, and will subsequently implement the actions when directed by the human. Where the human is responsible for making the decisions at Level 1, the automation proposes an optimal solution at Level 2, which the human can then either accept or amend.

Starting with Level 3, automation has sufficient autonomy and authority to initiate some of the actions without human involvement, which is further expanded at Level 4. The stepwise transition towards full automation with no human involvement (Level 5) will require mixed forms of operation, where parts of the system and/or traffic are increasingly being automated.

1

Table 1.1: Levels of automation from the European ATM Master Plan 2025, adapted from SESAR Joint Undertaking (2024, p. 68).

Level of automation	Tasks: P	A	D	E	Authority	
0 Low automation Automation gathers and exchanges data. It analyses and prepares all available information for the human operator. The human operator takes all decisions and implements them (with or without execution support).	A	A	H	H/A		N/A
1 Decision support Automation supports the human operator in action selection by providing a solution space and/or multiple options. The human operator implements the actions (with or without execution support).	A	A	H/A	H/A		N/A
2 Resolution support Automation proposes the optimal solution in the solution space. The human operator validates the optimal solution or comes up with a different solution. Automation implements the actions when due and if safe. Automation acts under human direction.	A	A	H/A	A		N/A
3 Conditional automation Automation selects the optimal solution and implements the respective actions when due and if safe. The human operator supervises automation and overrides or improves the decisions that are not deemed appropriate. Automation acts under human supervision.	A	A	A	A		
4 Confined automation Automation takes all decisions and implements all actions silently within the confines of a predefined scope. Automation requests the human operator to supervise its operation if outside the predefined scope. Any human intervention results in a reversion to Level 3. Automation acts under human safeguarding.	A	A	A	A		
5 Full automation There is no human operator. Automation acts without human supervision or safeguarding.	A	A	A	A	N/A	

Perception, **A**nalysis, **D**ecision and **E**xecution
Human and **A**utomation

Operations at Level 4 (confined automation) is the scope of this thesis. At this level, one sector may, for example, already operate at a higher LOA than its neighboring sector. Or certain aircraft may have capabilities on board, such as advanced navigation and communication aids or coupling with the autopilot (Sgorcea et al., 2016), that enable these aircraft to receive ATC services at a higher LOA[1] than other aircraft lacking this equipment. Even at high LOA, humans are expected to play an important role in supervising these future systems and to intervene when automation falls short (Metzger and Parasuraman, 2005); people will ultimately remain responsible. The automation will alert the ATCO to supervise (and potentially step in) when it detects that it needs to operate outside its predetermined scope. Level 5 (*"There is no human operator. Automation acts without human supervision or safeguarding"*, SESAR Joint Undertaking, 2020, p. 24) was only envisioned for future unmanned vehicle operations known as U-space, but has been excluded altogether from the 2025 edition of the master plan (SESAR Joint Undertaking, 2024).

[1]Note that whenever this thesis mentions LOAs it refers to the automation levels of the (next generation) ATC services, rather than that of the (current generation) aircraft.

1.3 Automation challenges

In many transportation domains, the technology is capable of handling the majority of situations. It is the remaining fraction that is difficult to automate in a safe and reliable way (Norman, 2015). The easy way out is to make human operators (whether a car driver, pilot or ATCO) responsible for intervening when the system cannot handle a situation. Unfortunately this has several significant drawbacks, referred to as 'ironies of automation' by Bainbridge (1983) that remain largely unresolved up to this day (Strauch, 2018). The increased popularity of artificial intelligence only aggravates this by introducing additional ironies (Endsley, 2023). Three ironies are particularly relevant for this thesis.

Firstly, humans are known to be bad at intervening in a task that they have not been actively involved in. Think of a self-driving car 'driver' who was not paying attention to the road and suddenly has to grab the steering wheel to avoid an accident (Nordhoff et al., 2023). As Hancock (2014, p. 453) puts it: *"if you build systems where people are rarely required to respond, they will rarely respond when required"*. It are precisely these difficult situations and tasks involving (sudden) high workload when the human would like to be supported by automation (Billings, 1996). SESAR explicitly states that ATCOs supervising automated systems require support tools to be able to intervene in unexpected events (SESAR Joint Undertaking, 2020, p. 87), reiterating the opinion of ATCOs themselves (Beadle, 2007).

Secondly, the importance of keeping operators engaged has been recognized (Chiappe et al., 2012) and is known to improve failure detection (Pop et al., 2012), situation awareness (Endsley and Kiris, 1995), motivation and job satisfaction (Endsley, 2017). ATCOs that are satisfied with their job are, in turn, more willing to accept new forms of automation (Bekier et al., 2011).

Thirdly, when human operators rarely perform a task, their (cognitive) skills in performing this task can deteriorate. During the COVID-19 pandemic, ATCOs reported skill erosion due to historically low traffic demands and subsequently had to spend more time in the simulator to stay proficient under high traffic loads (Kenny and Li, 2022).

Although listing these challenges separately may imply that they are standalone, in practice they are all intermingled (Edwards et al., 2017). Lowering workload, for example, can have a detrimental effect on situation awareness. Dekker (2004) compares this to an inflatable mattress, where pushing the air down in one part inevitably leads to a rise of air in another.

1.4 Towards human-automation teamwork

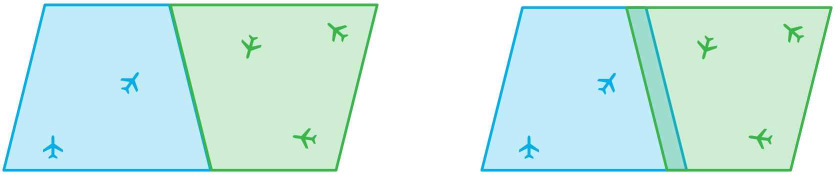
In highly automated systems, critical moments often occur after periods with little human involvement followed by a sudden need for the human to intervene. This intervention is made difficult when the automation has largely been working independently from the human, creating a disconnect between the two agents. Numerous studies suggest that humans (including ATCOs) should be kept in the loop by establishing teamwork between the human and automation (Endsley, 2017, Martin et al., 2016, Metzger and Parasuraman, 2001). If this is not done, the work domains of the two agents barely touch each other (Figure 1.3a) with only minimal exchange of information, a clear example of a potentially brittle team that would be unable to successfully handle situations when the automation's capabilities are inadequate (Klein et al., 2004).

(a) Lack of teamwork. (b) Teamwork established (overlap).

Figure 1.3: Schematic overview of human-automation teamwork.

Despite the fact that much ATCO work is currently performed as a human-human team effort (Soraji et al., 2012), most of the automation solutions proposed or implemented thus far are aimed at supporting individual ATCOs (Corver and Aneziris, 2015, Eurocontrol, 2024b, Maynard and Rantanen, 2005, Zon et al., 2009). In the sociotechnical community, there is an increasing belief that applying human-human teamwork concepts in human-automation teams (HAT) is the way forward (Endsley, 2017, Lyons et al., 2021). By increasing the exchange of information between both agents to establish a common frame of reference (Hoc and Carlier, 2002), the overlap is increased (Figure 1.3b). While back-up behavior is part of teamwork (Salas et al., 2005), the overlap should not be too large in order to prevent adding a significant amount of workload solely to establish and maintain teamwork.

To assign work to either agent, the Venn diagram from Figure 1.3 cannot only be applied to the agents themselves, but can also be applied to the physical world that the agents are acting in. Starting with the traditional sector-based allocation in Figure 1.2, the diagram can represent entire airspace sectors, with one sector fully automated and the other controlled by a human ATCO (Figure 1.4a). Then, the boundary between the two sectors will lead to a certain overlap between the agents in both sectors (Figure 1.4b) to ensure a streamlined transfer at the sector boundaries. In current-day operations, comprehensive letters of agreement are established between adjacent units (Nolan, 2017). These reduce the number of locations where flights interact across sector boundaries or shortly after entering a new unit, thereby requiring less cognitive workload than an airspace with many possible conflict locations. However, the reliance on designated coordination points (COP), altitude ranges, and additional separation criteria opposes many of the benefits seen in flight-centric ATC (FCA).



(a) Minimal sector interaction. (b) Sector interaction (overlap).

Figure 1.4: Schematic overview of sector interactions.

At a more fine-grained scale, the diagrams could represent individual flights. As long as there is no interaction between these flights, for instance when they fly in different sectors or directions, they can more or less be considered as isolated entities (Figure 1.5a). As soon as the flights have conflicting trajectories or trajectories that limit the other flight's solution space, they have a certain overlap (Figure 1.5b). If the flights are then handled by different agents (i.e., human and automation), it becomes necessary for these agents to create a similar overlap in their work domains in order to establish efficient teamwork. Such mixed conflicts are one of the key challenges of FCA, where it is proposed to have a 'less impacted flight' algorithm determine which ATCO should take action to resolve the conflict with minimal deviation from the planned trajectories (Finck et al., 2024).

(a) Flights with no interaction.

(b) Flight interaction (overlap).

Figure 1.5: Schematic overview of flight interactions.

If both flights are instead assumed by a single ATCO who delegated one of the flights to automation, it might be undesirable to assign the resolution of these mixed conflicts in the same way as in FCA. Unlike FCA's human-human setting, a human-automation team has larger asymmetry between the two agents. On the one hand, humans may prefer to always solve these conflicts themselves, as they can then stick to their own plan, a plan the automation is likely to be unaware of. On the other hand, humans are likely to have better situation awareness with respect to flights that are under their active control as compared to delegated flights, which could complicate and/or increase the time and effort required to manually resolve mixed conflicts.

Apart from these two extremes (allocation at the individual flight level or at the sector level), an intermediate form could consist of only delegating flights in a certain stream of traffic to either automation or human, as done by Finck et al. (2023b). Overflights, for example, could be delegated to automation, as they are generally less susceptible to conflicts than climbing and descending flights that need to be merged with other traffic. When automation is assumed to be capable of handling certain areas of the airspace or specific flights, there will still be some interaction between these flights and those controlled by a human.

To support the human in understanding, supervising and/or assessing the automation's actions in these interactions, systems can be introduced to increase automation transparency (Jans et al., 2019). Interacting with these systems inevitably comes with substantial mental demands for its operator (Wright et al., 2016). By tuning the allocation, the overlap between the two agents can be adjusted, which manifests itself in changes in human supervisory control performance in terms of attention allocation, workload, situation awareness, etc.

A concept along these lines is currently under development at MUAC where 'simple' flights are envisioned to be allocated to an automated system while 'complex' flights are still handled by human ATCOs (Hendrickx and Tisza, 2019). By handling flights at two different LOAs, it is expected that low vigilance, skill erosion and low engagement problems will be evaded, while simultaneously lowering the operator workload. This vision forms the main inspiration for this thesis.

1.5 Goals and contributions of this thesis

This thesis discusses the feasibility of modifying the allocation of flights to either human or automation in such a way that, whenever the human is required to act, he or she already has the related flights internalized in their mental model. Determining the impact of different levels of integration (overlap) and the role of flight allocation in those scenarios is at the core of this thesis. The smart allocation of flights can ease the formation of an efficient human-automation team, manifesting in improved human supervisory control performance and minimized human effort in handling mixed conflicts (Joe et al., 2014). The main research goal is therefore:

Research goal

Establish how flights can best be distributed between a human ATCO and an automated system, sharing control of an en-route sector, such that interference between the two agents/entities is minimized.

At first, it is important to get a good understanding of the current and future state of automation in the ATC domain, and what strategies different stakeholders apply in transitioning there. Knowledge from other domains aiming for higher levels of automation, such as the automotive industry, will also serve as an inspiration for the thesis. The literature survey is thus driven by:

Research question 1

How does the human-automation allocation of flights fit within existing strategies towards full automation in the ATC domain?

Previous works suggest that ATCOs are reluctant to adopt new forms of automation at the decision making level in their work (Bekier et al., 2012, Langford et al., 2022, Westin et al., 2016a). With most of the existing research focusing on automation at a functional level, it is not yet known whether sharing control over part of the traffic in an airspace would be a workable and acceptable situation from an ATCO stance. Therefore, the first exploratory research question is formulated as:

Research question 2

To what extent is the transfer of control of flights to an automated system dependent on system-proposed allocations, individual ATCO preferences and automation capabilities?

When flights are delegated to automation, they become 'foreign objects' within the airspace. The execution of certain tasks (mainly conflict detection and resolution) with respect to human-controlled flights will inevitably change. As ATCOs will keep the final authority over all flights for the foreseeable future, it might be beneficial to not delegate all tasks associated with a flight completely to automation, but leave some tasks with the ATCO. In doing so, the distribution of responsibilities, however, has the potential to lead to ambiguous situations. Understanding these changing tasks and their impact on the overall work of the ATCO feeds the third research question:

Research question 3

To what extent is the workflow of ATCOs affected by flights delegated to an automated system that interact with flights under their responsibility?

One of the most difficult situations in a shared airspace is the case of a mixed conflict, where the flights involved are controlled by different agents. The present-day operation relies on clear agreements on who is responsible for a certain part of airspace and tries to minimize the chance on inter-controller conflicts through the use of standard routes and procedures as set out in letters of agreement between adjacent ATC providers (Baumgartner, 2007). It might be beneficial to allocate flights such that the impact of these mixed conflicts is minimized. The classification of flights based on their individual complexity, which in turn depends on their (potential) interaction with other flights, serves as an input for determining this allocation. Whereas existing complexity metrics primarily focus on entire sectors (Hilburn, 2004, Prandini et al., 2011, Prevot and Lee, 2011), i.e., for personnel scheduling purposes, the complexity of individual flights is a less well-researched area.

Research question 4

Which other flights in the airspace add to the perceived complexity of an individual flight and what characterizes them?

After answering all the aforementioned questions, it remains to be seen if allocating flights based on their interaction is indeed favorable over allocations based on simpler rules, like the destination airport or sector exit. For this, a full-scale simulator experiment involving realistic traffic scenarios, where all elements from the preceding chapters come together, serves to answer the final research question:

Research question 5

Given a realistic traffic scenario, how should flights be allocated to either the human or automation, such that interactions between human- and automation-controlled flights are reduced, combined team performance is best supported and ATCO acceptance is increased?

1.6 Research scope

Given the broad topic of human-automation teamwork, the research has been narrowed down to match the available resources. First and foremost, the operational context is en-route ATC, as conducted at MUAC. Due to the less time-critical and more predictable nature of en-route ATC, high LOAs are expected to be adopted here first. Second, the focus is on allocating flights to either human or automation and its impact on one particular aspect of teamwork (Salas et al., 2005): working towards a common goal (the safe and efficient handling of air traffic). Finally, the impact on human supervisory control performance is considered in relation to full automation at the decision-selection and action-implementation level.

Additional assumptions are listed as follows:

Automation While automation is an important element in this thesis, the development of such automation is not part of it. Simple, rule-based automated solvers are used to provide a basic level of automation that is both predictable and reliable, and foremost easy to understand by the ATCO. The automated agent mimics the way an ATCO thinks (i.e., conformal automation, Westin, 2017), by following their decision making process and standard rules, such as keeping aircraft as high as possible and on direct routes. To increase predictability and in anticipation of future trajectory-based operations, the automation cannot issue heading or speed clearances and solves conflicts only by altitude.

The automation is fully capable of acting within the experiment scenarios without any human involvement, automatically executing actions to ensure safe and efficient air traffic.

Automation failures are outside the scope of this research. Together with the rule-based solvers, this diminishes or even removes the requirement for the highly independent automation to extensively communicate its intentions to the human operator, for which (complex) interfaces would be needed. The design of such interfaces is a research topic on its own, while this thesis focuses on the impact of flight allocation on teamwork, rather than the impact of inter-agent communication. If some level of communication is required for the sake of the experiment, present-day tooling is used as much as possible.

Participants and teamwork All experiments are performed with operational en-route ATCOs from MUAC to ensure realism and data validity. The participants are responsible for a certain airspace, optionally in co-operation with an automated agent. In current-day operations, ATCOs often control a sector in co-operation with a second ATCO, who is responsible for the coordination with adjacent sectors. This so-called dyad has been abandoned in this thesis, as the focus is on the teamwork between human and automation, rather than human-human co-operation. Future development, such as Single Controller Ops, may make this a feasible scenario (Gerdes et al., 2022). Coordination with adjacent sectors or pilots is for the same reason also neglected.

Teamwork, whether human-human or human-automation, is a broad topic involving many different aspects, such as leadership and communication (Salas et al., 2005). The focus of this thesis lies on one aspect: working towards a common goal.

Control task ATCOs are responsible for the safe and efficient handling of air traffic. Since speed clearances are given less frequently in en-route control, due to the narrow speed envelopes of aircraft flying at high altitude, aircraft are assumed to either be flying constant Mach throughout the simulations or follow a standard speed profile. ATCOs can therefore only issue heading, route and altitude clearances. This reduces the number of control strategies to resolve conflicts and further simplifies the automated solvers.

Note that, whenever this thesis speaks of 'control', it refers to the handling of flights by ATCOs in terms of issuing instructions to pilots, rather than the on-board control of aircraft by (auto)pilots. In addition, the final responsibility over all flights remains with the ATCO.

Infrastructure Controller pilot data link communications (CPDLC) is an essential element for future trajectory-based operations and is already increasingly in use today (Alharasees et al., 2022). Future automation is not expected to communicate via voice at all. This also provides human and automation a level playing field in terms of ground-air communication, information acquisition and uplinking clearances to aircraft. This thesis assumes CPDLC to be implemented, all other forms of communication with pilots are neglected.

Simulation The simulation environment is designed to mimic the current-day MUAC human-machine interface to provide high face validity (Dow and Histon, 2014) and reduce the required training/familiarization time. To this end, only support systems that are already in use at MUAC, such as the verification and resolution advisory tool (VERA), are made available to the ATCOs. Furthermore, Base of Aircraft Data (BADA) 3.10 performance models from Eurocontrol (2012) are used to provide realistic aircraft behavior, although reference weights are used for each aircraft type. The aircraft are flown by scripted autopilots that always adhere to instructions issued by the ATCO or automation. No human pseudo-pilots are used.

Data quality Uncertainties and unpredictability such as variable delays in pilot responses and changing weather are eliminated, ensuring perfect automation solutions. In conjunction with the absence of automation failures, this removes the considerable impact that (a lack of) trust in automation might have, which is beyond the scope.

1.7 Thesis outline

This thesis consists of three parts relating to three phases of research. Figure 1.6 shows how the various chapters connect. Each chapter addresses one of the research questions stated in Section 1.5.

Phase I contains the exploratory phase of the research.

Chapter 2 provides an overview of existing literature on automation, as well as specifics in an ATC setting. It also introduces the state of the art at MUAC and their envisioned future concept of operations with respect to automation. It serves as a foundation for the thesis.

Chapter 3 discusses results of a preliminary experiment performed at MUAC to provide initial insights in the concept of sharing traffic between a human and an automated system in a single airspace, and to examine the willingness of ATCOs to delegate flights. Lessons learned from this experiment have been used as a basis and inspiration for Chapters 4 and 5. Also, the MUAC-style simulation platform used throughout the thesis was first tested in this experiment.

Phase II takes a deeper dive into what makes specific flights more complex for ATCOs than others and how to model or predict this complexity.

Chapter 4 describes the development of a control task and strategy framework in the form of cognition flowcharts for conflict detection and resolution. The charts are based on observations in the experiment of Chapter 3 and in MUAC's operations room. The charts show connections between these typical ATCO (sub)tasks, together with empirically collected time trace data to quantify the complexity of these tasks, and the potential impact of delegating flights to automation on this.

Chapter 5 discusses the perceived complexity of individual flights, in relation to surrounding flights. The resulting complexity metric can be used to drive an automated flight allocation algorithm.

Phase III puts the insights from Phase II into practice, reflecting on their practical use as basis for an allocation algorithm.

Chapter 6 applies the lessons learned in testing two allocation schemes in a full-scale experiment with a relatively large number of ATCOs and a realistic traffic scenario. Compared to the experiment from Chapter 3, here the allocation is further scrutinized in a more controlled environment.

The findings from this thesis are discussed and concluded in **Chapter 7**, together with recommendations for further research and potential operational use of individual flight allocation.

Appendix A introduces the SectorX simulator that was used in all experiments in this thesis, as well as several experiments from other research projects. It also provides pseudo-code of the automation solvers used in Chapters 3 and 6.

Appendix B contains the experiment briefing and questionnaire of the experiment from Chapter 6. These are exemplary for the other experiments.

Appendix C provides a literature survey of techniques often used in ATC research to mitigate recognition of repetitive scenarios. Using data from the experiment in Chapter 4, it discusses the delicate balance between preventing recognition and simultaneously measuring conditions that are – except for the studied independent variable(s) – as much alike as possible.

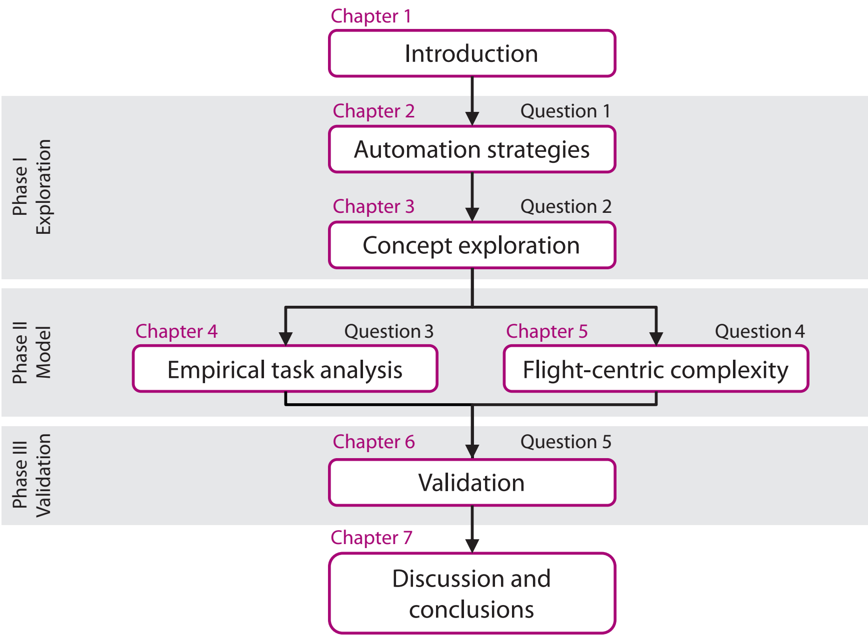


Figure 1.6: Thesis outline.

2

Automation strategies

As seen in Chapter 1, many domains, including air traffic control (ATC), are moving towards higher levels of automation to increase capacity, safety and efficiency, while coping with predicted staff shortages. After a short introduction to automation in general as well as several key concepts, this chapter serves as a basis for the thesis, by setting the stage for future automation developments in ATC. Two automation strategies, or paths, towards full automation are discussed in an ATC context. The chapter furthermore introduces ongoing automation projects in the ATC domain, with a special focus on Maastricht Upper Area Control Centre (MUAC), whose ARGOS project is closely related to the work in this thesis.

Parts of this chapter have been published in:

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

2.1 Introduction to automation

Automation can be defined as *"the execution by a machine agent (usually a computer) of a function that was previously carried out by a human"* (Parasuraman and Riley, 1997, p. 231). In the traditional *technology-centered* designs, automation was indeed used to relieve human operators by taking over (some of) their tasks (Endsley and Kaber, 1999). This effectively moves the human away from the control loop into a supervisory role, monitoring the automation.

This significant shift from the use of 'dumb' (analog) tools, where humans have full control over all aspects (Sheridan and Parasuraman, 2005), brings forward a number of issues as articulated by Bainbridge (1983) that are still valid today (Strauch, 2018) and remain valid in the future with the growing use of artificial intelligence (AI, Endsley, 2023). Operators can, for example, wrongfully assume that the automated system they are supervising is functioning nominally, leading to complacency (Wickens et al., 2015). This form of non-calibrated trust can also occur the other way around, with operators distrusting the automation and therefore ignoring its alerts (Parasuraman et al., 2000). Furthermore, operators are at risk of losing their vigilance, manual skills and situation awareness when the system is performing most of the work.

Advances in computing power and more reliable technology are enabling the development of smarter, more autonomous agents that can make independent decisions and adapt to circumstances (Rieth and Hagemann, 2022). Automation is no longer limited to performing pre-programmed, fixed tasks that a human has asked it to perform, but can perform tasks without external (i.e., human) involvement. Eventually, this may evolve into a fully autonomous system where no human involvement is required at all.

The onset of autonomy comes with new challenges, where automation should no longer be seen as a tool but as a nearly equal collaborator (De Visser et al., 2018). This can be attained by applying a *human-centered* approach in which automation supports or complements the operator (Billings, 1997). Not taking away all authority from humans is known to help them gain trust in a system (De Visser et al., 2018) and keeps them involved and skilled. Determining an adequate function allocation is a delicate undertaking though, requiring careful consideration (Pritchett et al., 2014).

2.2 Automation in ATC

2.2.1 Present-day automation

Recalling Figure 1.1, it is clear that the air traffic control (ATC) domain is not adverse to automation. In fact, present-day controller working positions (CWP), as shown in Figure 1.1e, widely embrace automation support in the display and processing of information. Flight plans, downlinked aircraft information and air traffic control officer (ATCO) inputs are automatically associated with the corresponding radar plots and populate the related aircraft labels. Another accepted use of automation is in the conflict scanning and alerting task. Alarms go off when two aircraft are predicted to lose separation if adhering to their cleared trajectories, and warnings are displayed when an alternative trajectory (desired by the ATCO) would result in a loss of separation. The detection horizon varies from circa two minutes for short-term conflict alerting (STCA) to up to 20 minutes for medium-term conflict detection (MTCD) (Eurocontrol, 2007, 2017). MTCD makes use of a flight's predicted trajectory based on its flight plan, whereas STCA simply takes the current speed vector (position, heading and speed) to determine a future position. The introduction of 4D trajectory-based operations is expected to make MTCD predictions more reliable in the coming decades, although some intrinsic uncertainty will always remain (Paglione et al., 2017).

More recent developments include the adoption of probing tools that can show the feasibility of an intended clearance. At Maastricht Upper Area Control Centre (MUAC), this is implemented by the highlighting of unsafe control actions in the clearance menus and, upon request, the highlighting of flights that conflict with the intended clearance (Eurocontrol, 2024b). This is a first step towards more support in the resolution task. Even though various algorithms have been developed that can automatically provide solutions to solve conflicts (Alaeddini et al., 2011, Erzberger et al., 2012, Frazzoli et al., 2001, Gariel and Feron, 2009), the inability of proving that these are always safe severely complicates certification for operational use. They are therefore not yet in service, leaving the resolution task to the human as creative and adaptive decision-maker.

Unlike the decision stage, where a solution needs to be chosen from a set of options, executing (or implementing) the selected option is already more automated (Hendrickx and Tisza, 2019), as seen with the automated uplinking of clearances for example. Human decision making is for good reasons one of the most, if not the most, difficult stages to automate, as widely recognized in literature (Endsley and Kaber, 1999). It is not without reason that ATCOs around the world see automation as a helpful and useful addition, as long as it supports them instead of taking over their decision making (Bekier et al., 2012, Langford et al., 2022). ATCOs generally prefer to be 'in control'. Apart from technical challenges, legal constraints currently prohibit the use of automated decision making and execution, as the executive ATCO is ultimately still responsible (Contissa et al., 2012).

2.2.2 **Future visions on automation**

In its 2025 master plan for the European ATC system, SESAR Joint Undertaking (2024, p. 68) presents an automation roadmap towards higher levels of automation (LOAs), as shown in Table 2.1. This automation taxonomy is based on a levels of automated driving standard (SAE International, 2021), in combination with four stages (or tasks) devised by Parasuraman et al. (2000).

In Levels 0 and 1, a human operator is still fully responsible for executing all actions, whether they are self-created or proposed by the system. The current European ATC system is at Level 0, and progressing towards Level 1 with limited action selection support. At Level 2, expected by 2035, the automation proposes the optimal solution for the ATCO to validate or reject. The automation can now implement actions when approved by the ATCO who assumes a 'director' role (i.e., management by consent, Billings, 1997).

From Level 3 onward, the automated system receives an increasing amount of authority, which means that it can autonomously execute actions unless the ATCO intervenes (i.e., management by exception, Billings, 1997). Here the human acts as a supervisor. At Level 4, ATCOs will no longer directly guide traffic but will act as 'safeguarder'. They will only be asked to supervise when a situation is outside the automation's predefined scope and potentially requires their intervention, which would revert the system to Level 3. Compared to the previous edition of the master plan (SESAR Joint Undertaking, 2020, p. 24), 'High automation' at Level 4 has been rephrased to 'Confined automation' in

Table 2.1: Levels of automation from the European ATM Master Plan 2025, adapted from SESAR Joint Undertaking (2024, p. 68).

Level of automation	Tasks P	Tasks A	Tasks D	Tasks E	Authority	
0 Low automation Automation gathers and exchanges data. It analyses and prepares all available information for the human operator. The human operator takes all decisions and implements them (with or without execution support).	A	A	H	H/A		N/A
1 Decision support Automation supports the human operator in action selection by providing a solution space and/or multiple options. The human operator implements the actions (with or without execution support).	A	A	H/A	H/A		N/A
2 Resolution support Automation proposes the optimal solution in the solution space. The human operator validates the optimal solution or comes up with a different solution. Automation implements the actions when due and if safe. Automation acts under human direction.	A	A	H/A	A		N/A
3 Conditional automation Automation selects the optimal solution and implements the respective actions when due and if safe. The human operator supervises automation and overrides or improves the decisions that are not deemed appropriate. Automation acts under human supervision.	A	A	A	A		
4 Confined automation Automation takes all decisions and implements all actions silently within the confines of a predefined scope. Automation requests the human operator to supervise its operation if outside the predefined scope. Any human intervention results in a reversion to Level 3. Automation acts under human safeguarding.	A	A	A	A		
5 Full automation There is no human operator. Automation acts without human supervision or safeguarding.	A	A	A	A	N/A	

Perception, **A**nalysis, **D**ecision and **E**xecution
Human and **A**utomation

the 2025 edition, acknowledging the belief that high LOA may be applicable in a limited scope, rather than trying to handle all situations at a single automation level.

At Level 5, 'Full automation', the system operates completely without the ATCO. This level has been excluded altogether from the 2025 edition, but is kept here to indicate the highest theoretical level. SESAR aims to reach Level 4 by 2045. The table clearly shows that SESAR postpones the transfer of full decision and action selection authority to the automation. This is indicative for the widely supported belief that, ultimately, ATCOs remain and should be in control for the foreseeable future (Bekier et al., 2012).

2.2.3 Enablers for future ATC

A plethora of research projects have been and still are devoted to researching and developing novel technologies that can enable future ATC systems (SESAR Joint Undertaking, 2019). Trajectory-based operations (TBO) are a prominent example that will have flights follow pre-negotiated 4D trajectories (position and time) that are constantly updated by all stakeholders to provide accurate predictions (Enea and Porretta, 2012). These trajectories can then be cleared of conflicts, even before a flight departs (Pérez-Castán et al., 2020). However, intrinsic uncertainties like unforeseen events or changing weather conditions remain and can prevent flights from maintaining the negotiated trajectories (Corver and Grote, 2016), necessitating real-time ATC.

The downlinking of extended projected profiles (EPP) from the aircraft's on-board flight management system (FMS) through automatic dependent surveillance - contract (ADS-C) will greatly enhance the accuracy of ground-based vertical and speed profiles (Haugg et al., 2015). By sharing the ideal (from the aircraft's point of view) top of descend, ATCOs can issue optimized descent clearances and subsequently save fuel. As of 2024, circa 130 of the 5,500 daily flights at MUAC establish an ADS-C connection, which is expected to increase to 800 flights in 2026, in anticipation of a mandatory regulation for all newly-built European aircraft in 2028 (Jagasits, 2024).

An additional prerequisite for automated controllers is an increased adoption of controller pilot data link communications (CPDLC), as already mandated for nearly all flights operating above FL285 in European airspace since 2020 (Alharasees et al., 2022). This is fundamental as next to, e.g., reducing radio congestion and improving transmission clarity, it removes the dependence on the human voice for air-ground communication. Several instructions can be issued in parallel, which was not possible with radio only. While talking to one flight, the ATCO can uplink a few more instructions to some other flights – which is a significant boost to ATCO productivity. In a similar fashion, CPDLC will allow a human-automation team to instruct several flights in tandem – without disrupting the workflow of one another.

Despite all its benefits, CPDLC has several important shortcomings as well, such as increased latency compared to radio and reduced situation awareness of pilots (i.e., no 'party line' where pilots can hear instructions to other flights, Etherington et al., 2019), that can hinder the introduction and acceptance of automatic controllers. In fact, CPDLC is currently not to be used for short-term time-critical instructions (International Civil Aviation Organization, 2017), where voice is expected to remain the prevalent means of communication due to its intrinsic speed and instant feedback/acknowledgment. However, it are precisely flights requiring short-term tactical control that are expected to remain the responsibility of an ATCO in a shared human-automation airspace.

In tandem with the developments of improved trajectory prediction and data exchange infrastructures, rapid advances in AI-based technology are paving the way for the creation of 'digital ATCOs' to safely and efficiently direct flights through the airspace (Ortner et al., 2022). Examples include experimental data-driven approaches to conflict detection (Pérez-Castán et al., 2022) and conflict resolution (Pham et al., 2022) that can augment well-established physics-based models and algorithms. Currently, legal barriers and certification hurdles prohibit the use of (nondeterministic) AI for automated decision making and execution, leaving the human ATCO with the ultimate responsibility (Contissa et al., 2012, Lanzi et al., 2021). Therefore, it is expected that advanced data-driven and physics-based algorithms will initially be integrated into decision-support tools to *assist* rather than replace the ATCO.

Indeed, many of the ongoing projects that expand the role of automation still rely on some form of human involvement. Single-controller operations, for example, transition from the current human-human ATCO dyad to an automation-human dyad (Gerdes et al., 2022, Hunger et al., 2024). Fully automated ATC is still far on the horizon, but that does not preclude the introduction of new forms and higher LOAs in the forthcoming decades.

2.3 Strategies towards full automation

The strive towards full automation in the ATC domain has interesting parallels with the automotive industry, whose automation strategies can serve as a source of inspiration for ATC. Cars increasingly receive automation support functions. For instance, cruise control is a standard feature on most modern cars and allows the human driver to set a desired speed that the vehicle then maintains. As more automation is introduced, cars start making more autonomous decisions without human involvement. One such feature is adaptive cruise control, which automatically slows the vehicle when it is approaching another vehicle in front. Nevertheless, the human driver still needs to steer the car. When a driver wishes to overtake the preceding vehicle, the system cannot reliably anticipate exactly when the driver will steer towards the adjacent lane (which can vary with driver preference) and may therefore start braking, which the driver did not anticipate. This lack of mutual understanding is just one example of many potential human-automation interaction issues.

On the one hand, such a gradual *function*-based introduction to more automation increases acceptance by allowing operators to slowly adapt and gain trust and confidence in the automation (Cioroaica et al., 2020). Incremental automation can make operators feel like they remain in control, while delegating fine-grained control to automation. On the other hand, this approach slows down achieving higher levels of automation, as it requires solving many intermediate mixed-authority challenges that do not directly add to the development of full automation.

As an alternative, systems that operate within a *constrained* environment, such as trains or shuttle buses on dedicated lanes, can overcome many of these issues by operating at very high LOAs or even fully autonomously (Schutte, 2017, Stayton and Stilgoe, 2020). Proving that automation can work under any condition and in any environment (i.e., unconstrained) is no easy feat though. It is therefore that the most advanced autonomous cars are currently operating at Level 4 at most on an automotive industry-standard defined by SAE International (2021). These vehicles can only operate within geofenced areas (e.g., city centers or highways) and under specific (weather) conditions, and require human intervention when exiting the constraints. Similar developments are seen in the maritime domain with remotely supervised autonomous ships (Rødseth, 2021). For Level 5, the constraints need to be lifted such that the vehicle can operate anywhere at anytime.

2.3.1 Comparing strategies

To compare the aforementioned function- and constraint-based strategies in an ATC context, we introduce the chart in Figure 2.1. The numbered boxes denote the respective system level (referring to the system as a whole), while the boxes' vertical and horizontal location indicate the share of flights in an airspace that operates at a specific LOA. Since present-day ATC makes heavy use of automation in the information acquisition and analysis stages (e.g., conflict alerts and label correlation), most LOA taxonomies used in industry do not start at a fully manual level. Despite the misleading name, Level 0 often does involve some form of automation, as reflected by the gray area on the left of the chart.

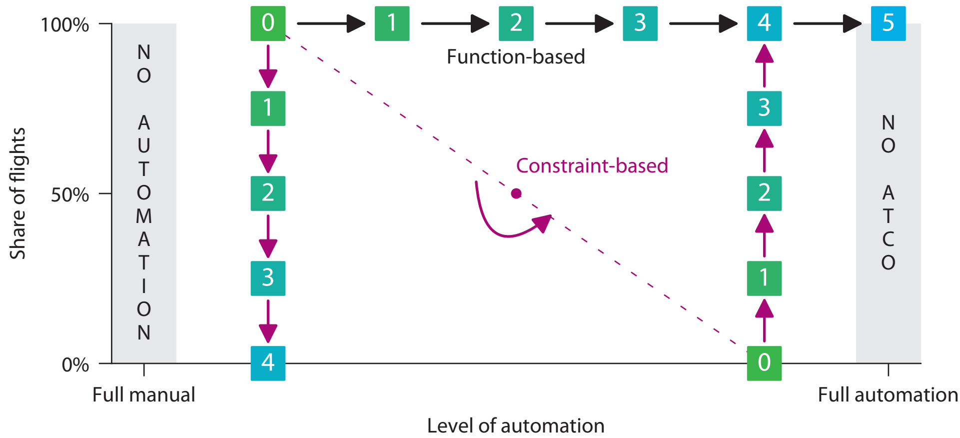


Figure 2.1: Function- and constraint-based automation strategies towards full automation.

At the highest level on the far right of the figure (Level 5 in this case), human ATCOs are no longer involved. This is not foreseen to be attainable in the medium-term future (or even at all) and therefore grayed out as well. Note that for both strategies the number of intermediate levels can vary and that the horizontal axis is an *ordinal* scale (i.e., the exact horizontal distance between levels does not have any meaning).

Function-based

Traditional function-based automation strategies can be seen as one-dimensional, when evaluated at system-level, with the LOA gradually, but system-wide, increasing as the capabilities of the automated system advance over time (i.e., moving from left to right in Figure 2.1). Examples of this can be found on the flight deck (e.g., the first autopilots could only maintain attitude and heading whereas modern systems can follow complete routes, Landry, 2009) and in the automotive industry, where Tesla has been gradually increasing the capabilities of its 'autopilot' systems (Kannan and Lasky, 2020).

Constraint-based

An alternative, constraint-based strategy, is to promptly attain a high(er) LOA in a constrained environment and gradually expand this environment. Looking at the automotive industry once again, companies like Waymo abandoned the development of SAE Level 3 systems in favor of Level 4 as successor to the widely available driver support systems at Level 2 (Kannan and Lasky, 2020). Their self-driving taxis can operate autonomously, but only in a constrained environment such as specific urban areas.

In Figure 2.1 this strategy pivots around a central point and has simultaneous – but counterbalanced – up and down movements at two distinct LOAs, enabling a higher LOA for certain flights only, before making this available to all flights. It can be thought of as a traditional balance scale, where an increase in flights operating at one LOA is accompanied by an equal reduction in flights operating at the other LOA.

2.3.2 Serial and parallel automation

Function-based automation generally invokes serialized interactions in which the human operator needs to monitor the automation and/or accept or reject solutions proposed by it (Endsley, 2017), as shown at the top of Figure 2.2. In such a setup, described by Millot and Mandiau (1995) as a vertical system, the human is mostly backing up the automation, leading to reiteration of a large part of the work. Serial automation at the decision-making and execution stages is often not efficient (Endsley, 2017), as operators may prefer to or even need to re-analyze a situation. For example, if ATCOs need to check whether a solution presented by the automation is feasible, they may perform a similar or even more time-consuming analysis compared to when they had to come up with the solution themselves. Similar to on-the-job training, where experienced ATCOs sit alongside trainees to monitor their actions, automation is degraded to a 'student' that requires close supervision. Therefore, the desired workload reduction is often not attained with serial automation in complex tasks (Endsley, 2017). In addition, serial processes require high levels of conformance with respect to the ATCO's individual style in order to increase acceptance (Hilburn et al., 2014, Westin et al., 2016b). Serial automation is often implemented as a hierarchy, with the human governing and delegating tasks to automation.

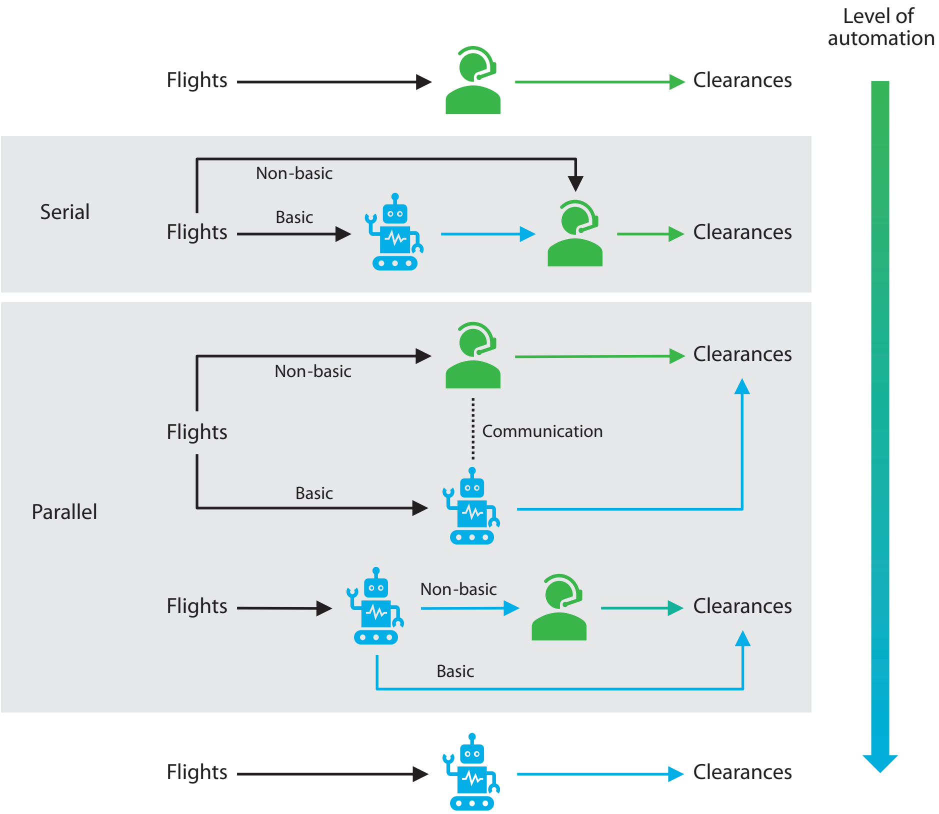


Figure 2.2: Serial versus parallel automation at various LOA, adapted from Endsley (2017). 'Basic' and 'Non-basic' refer to the complexity of individual flights' trajectories through the airspace.

Constraint-based automation leads to a parallel (or horizontal, Millot and Mandiau, 1995) system that evokes a heterarchy, that is, a level playing field where both human and automation can take initiative and execute actions (Pacaux-Lemoine and Flemisch, 2019, Rieth and Hagemann, 2022) as illustrated in the lower half of Figure 2.2. In contrast to a serial system, the automation now gets sufficient authority to make autonomous decisions that do not require human input. In ATC, the most extreme form of parallelism can be achieved by fully automating an entire sector while a neighboring sector remains completely under manual control.

In practice, flights in both sectors can have an impact on each other, meaning that true parallelism can only be approximated, as discussed in Chapter 1. A common frame of reference (Hoc and Carlier, 2002) or team orientation (Joe et al., 2014) is a prerequisite for efficient parallel systems, making joint communication equally or even more important (Feigh and Pritchett, 2014). Experienced human-human teams often rely on implicit communication, such as body language and voice nuances, which is naturally very difficult for computers (Joe et al., 2014). In current-day operations, with a human ATCO in every sector, mutual performance monitoring and backup behavior is provided by placing the ATCOs responsible for adjacent sectors side-by-side, or on direct phone lines with each other (Svensson et al., 2019). Insufficient communication can lead to compensating behavior and a (temporary) return to a serial system. The main challenge in creating a parallel system is, therefore, how to organize the overlap between the ATCO and automation work domains to ensure that not too many serial processes co-exist.

2.4 Function-based strategy

The traditional approach towards full automation is to gradually increase the LOA of specific functions (Figure 2.1). In such a function-based automation strategy, there is a continuum between 'none' and 'full' automation with many intermediate steps where part of a (sub)task is performed by a human and part by automation. Intermediate steps seem to have advantages over near-full automation with a human supervisor by (theoretically) enabling human-machine teamwork (Kaber and Endsley, 2004), but partial automation can also lead to increased operator workload compared to manual operation, as demonstrated in car driving experiments by McDonnell et al. (2023).

2.4.1 Stages and levels of automation

To describe the various intermediate stages of an automated system, LOA taxonomies like SESAR's (Table 2.1) are widely used in many variations (Vagia et al., 2016). A system can simultaneously operate at multiple LOAs by assigning a distinct level to specific tasks or situations, but a single LOA can also be used to describe the system as a whole.

In 1978, Sheridan and Verplank were the first to list ten LOAs on a one-dimensional scale for robot teleoperation (Table 2.2). At all levels the automation implements the job, but with increasing LOA, human involvement is reduced in exchange for more autonomy of the automated system. At Level 1 the human tells the automation what to do, at Levels 2–6 the automation suggests actions to the human and at Levels 7–10 the automation only informs the human about its decision after it has implemented it. As can be seen, the scale mainly refers to making a decision out of several options and subsequently implementing the chosen action, i.e., the *output* of a system.

Table 2.2: Levels of automation in human-computer decision-making, adapted from (Sheridan and Verplank, 1978, pp. 8-17–8-19).

LOA	Description
1	Human does the whole job up to the point of turning it over to the computer to implement
2	Computer helps by determining the options
3	Computer helps determine options and suggests one, which human need not follow
4	Computer selects action and human may or may not do it
5	Computer selects action and implements it if human approves
6	Computer selects action, informs human in plenty of time to stop it
7	Computer does whole job and necessarily tells human what it did
8	Computer does whole job and tells human what it did only if human explicitly asks
9	Computer does whole job and tells human what it did, if the computer decides he should be told
10	Computer does whole job if it decides it should be done, and if so tells human, if it decides he should be told

Parasuraman et al. (2000) later realized that automation can also play a role in the *input* preceding the decision-making. They therefore extended the taxonomy by formulating four consecutive stages reflecting human information processing: a) information acquisition, b) information analysis, c) decision and action selection, and d) action implementation. The first two and latter two stages relate to, respectively, the input and output of the system. Each stage has a specific LOA ladder with a variable number of steps, dependent on the system at hand.

The original LOA taxonomy discussed so far is rather generic and not easily applicable to cognitively demanding domains like ATC. Endsley and Kaber (1999) therefore came up with another ten-level taxonomy (Table 2.3) that describes in detail who performs what role at each level: the human operator, the automation, or a mixture of both. The associated four stages (or roles) are more in line with the typical tasks of an ATCO. As the LOA increases, automation involvement gradually expands to more action-based roles. At the same time, human involvement sees an opposite shift towards a more monitoring role. In the table this shows as the green area shrinking to the right with increasing LOA, while the blue area expands to the left. In contrast to, e.g., the taxonomy from SESAR (Table 2.1), here full action-implementation autonomy is allocated at an earlier stage. Although this taxonomy explicitly acknowledges that many functions are performed as a co-operation between human and automation, it does not further detail this co-operation, making it too superficial for system designers (Kaber, 2018).

In an attempt to tackle this, Save and Feuerberg (2012) developed a new LOA taxonomy for the European SESAR program, specifically aimed at designs for systems supporting flight crews and ATCOs (Table 2.4). Its number of LOAs varies per function (or stage), as was already suggested by Parasuraman et al. (2000): there are six levels for both perception and analysis, seven for decision and nine for execution. The higher number of

Table 2.3: Hierarchy of levels of automation, adapted from (Endsley and Kaber, 1999, p. 466).

Level of automation		Roles: Monitoring	Generating	Selecting	Implementing
1	Manual control	H	H	H	H
2	Action support	H/A	H	H	H/A
3	Batch processing	H/A	H	H	A
4	Shared control	H/A	H/A	H	H/A
5	Decision support	H/A	H/A	H	A
6	Blended decision making	H/A	H/A	H/A	A
7	Rigid system	H/A	A	H	A
8	Automated decision making	H/A	H/A	A	A
9	Supervisory control	H/A	A	A	A
10	Full automation	A	A	A	A

Human and **A**utomation

levels for the decision and execution tasks reflect the difficulty and sensitivity of automating these tasks that are key for autonomy. At Level 0, all tasks are manually performed, without any support, while at Level 1, *"the human is accomplishing a task with 'primitive' external support, which is not automation as such"* (Save and Feuerberg, 2012, p. 47). Note that from Levels 4 and 5 onward in, respectively, the decision and action functions, automation switches from purely supporting human decision making and action execution to actually making and executing the decision autonomously. In contrast, both information acquisition and analysis do not go beyond the 'support' level and leave room for human involvement, even at their most automated levels.

Despite their widespread use, LOA taxonomies have several shortcomings that have caught attention in recent years. Kaber (2018) suggested that the taxonomies should evolve to more descriptive human performance models. In response, Wickens (2018) proposed some solutions, for example by combining the information acquisition and analysis stages to a 'situation assessment' stage. Similarly, he argued that the decision making and action stages could be combined into a single 'choose and execute' stage. Other criticism is aimed at the apparent ordinal nature of numbered levels, while in reality it is a nominal scale, that is not necessarily listed in a particular order that needs to be traversed from top to bottom (Stayton and Stilgoe, 2020).

This also implies that the highest level is not a 'holy grail' that must be reached whatever the cost. This was already coined by Parasuraman et al. (2000), who suggested that the appropriate LOA depends on the risks associated to automating a certain function, but seems to have been replaced by an overoptimistic belief in computer capabilities over the years. For example, Martin et al. (2016) showed that introducing full automation in an ATC setting resulted in worse overall system performance than an intermediate level, where both human and automation were involved. System designers could skip LOAs or systems can revert back to lower levels for better performance. On the flight

Table 2.4: Levels of automation taxonomy, adapted from (Save and Feuerberg, 2012, pp. 48–50).

	Functions			
LOA	Information acquisition	Information analysis	Decision and action selection	Action implementation (execution)
0	H: without any tools	H: without any tools or support	H: generates and selects options	H: executes and controls actions manually
1	H: non-digital artifacts	H: non-digital artifacts	H: generates and selects options with non-digital artifacts	H: executes and controls actions with non-digital artifacts
2	A: supports acquisition H: integrates and filters	A: helps compare, combine and analyze on request	A: proposes options H: selects one option or requests new options	A: executes parts on request or provides execution guidance H: full execution control
3	A: supports acquisition, helps integrating and filtering based on user settings	A: helps compare, combine and analyze on request, triggers alerts H: request help	A: proposes option(s) H: selects one proposal or requests new options	A: executes on request H: initiates, modifies or interrupts
4	A: supports acquisition, integrates and filters based on predefined criteria (visible to user)	A: helps compare, combine and analyze, alerts user	A: generates options and decides H: always informed	A: executes on request H: initiates, monitors or interrupts
5	A: supports acquisition, integrates and filters based on predefined criteria (not visible to user)	A: compares and analyzes data based on predefined parameters, alerts user	A: generates options and decides H: informed on request	A: initiates and executes H: monitors, modifies or interrupts
6			A: decides H: never informed	A: initiates and executes H: monitors and interrupts
7				A: initiates and executes H: monitors partially, limited interruption opportunities
8	No higher levels defined by Save and Feuerberg			A: initiates and executes H: cannot monitor nor interrupt

Human and **A**utomation

deck, for example, it is a widely recognized practice to revert to a lower LOA in abnormal flight conditions (Moriarty, 2015). In the case of system failure or limitations, graceful degradation is preferable over catastrophic failure (i.e., reverting to a lower level, rather than disabling all automation) (Edwards and Lee, 2017).

Another important realization, already introduced in Chapter 1, is that higher LOAs may be easier to implement than lower levels as human involvement is reduced. Think of autonomously driving trains that have been operating for decades, albeit in constrained environments often physically separated from user-operated vehicles, such as on elevated tracks between airport terminals (SAE International, 2021).

2.4.2 Adaptable and adaptive automation

Depending on its design, a system can operate at a static LOA (as chosen by the system designer) or dynamically change to higher or lower levels in real time (Lagu and Landry, 2010). Many automated systems switch to a lower LOA when they detect a situation that they cannot handle, giving back some or even full control to their human operators. There can, however, also be human-centered reasons to (temporarily) lower the LOA. While a high LOA may lead to improved overall performance and lower operator workload, it can simultaneously lead to a reduction in situation awareness and operator skills (Metzger and Parasuraman, 2001, Onnasch et al., 2014, Parasuraman and Wickens, 2008). Humans are, for example, more capable of detecting changes when such changes are controlled by themselves rather than automation or other humans (Wickens, 1995). Essential skills, like conflict detection and resolution in ATC, should not be fully automated at all times but may need to be under certain conditions (Landry, 2012). Finding a method to cater for such a dynamic balance is key in developing a successful human-automation system (Calhoun, 2021).

The more LOAs available to a system, the smaller the difference between levels and the larger the potential for confusion about the active level, recognized as 'mode confusion' on the flight deck by Sarter and Woods (1995). As this risk is largely dependent on who gets to decide the active LOA, a further distinction can be made between adaptable and adaptive automation (Parasuraman and Wickens, 2008). *Adaptable* automation requires the user to initiate changes in presentation modes or function allocation, whereas in *adaptive* systems such changes are initiated by the system. *Hybrid* automation combines both forms, with user and system jointly capable of changing the LOA.

Of the two, adaptable automation is vastly easier to implement, as it assigns the allocation responsibility solely to the operator. This in itself may lead to a workload increase though (Bailey et al., 2006, Kirlik, 1993). A further complication is that humans are biased in assigning interesting or challenging tasks to themselves, even though automation may be better in performing those tasks (Hopkin, 1998).

With adaptive automation, a system could automatically switch to a different LOA after a detected change in e.g., an operator's decision-making quality (IJtsma et al., 2022) or mental workload (Aricò et al., 2016). In a study with professional ATCOs, Di Flumeri et al. (2019) showed that adaptive automation driven by electroencephalography (EEG) and eye tracking measures is indeed capable of increasing operator performance and satisfaction by re-evaluating the active LOA every 5 minutes. Their study was limited to only two LOAs though, meaning that the active LOA was often either too high or too low. Determining exactly when to change an adaptive system to a different level is a delicate task that has not yet been reliably solved in the ATC domain.

Both forms of automation have their advantages and disadvantages (Calhoun, 2021). In general, adaptive automation works best when applied to the information acquisition and action implementation stages, rather than the more cognitively demanding analysis and decision stages (Kaber et al., 2005). The Federal Aviation Administration, among others, recommends adaptable automation over a static LOA, as it leaves the operator actively in control (Ahlstrom, 2016). This was echoed by Rieth et al. (2024) in a recent study involving 126 professional ATCOs.

2.4.3 Advantages and disadvantages

The stepwise technology-centered expansion of automated functions allows for relatively easy tasks, such as transfer of control to the next sector, to be automated first. On the one hand, this allows operators to gradually gain trust in the automated functions, one at a time. On the other hand, this easily leads to many intermediate steps where automation does (an increasing) part of the job and the human does another part, which is prone to the aforementioned 'ironies of automation' (Section 1.3). Rather than full automation, it is generally the intermediate levels that are problematic (Norman, 2015).

For this reason, Young and Stanton (2023) propose that the automation of a function should preferably be delayed until the automated system is mature enough to act fully autonomously without human involvement. They dubbed this human-centered approach the 'cliff-edge' principle after the envisioned sudden jump in LOA compared to the traditional step-wise increments across intermediate LOA, as illustrated in Figure 2.3 where it circumvents the 'area to avoid'. Note that here the vertical axis depicts the level of 'human involvement', rather than the share of flights. Although the two may seem similar, human involvement can vary independently of the share of flights.

Because reaching mature automation is a lengthy process, especially in safety-critical domains like ATC, strictly following the cliff-edge principle would mean that no higher LOAs could be implemented in the coming decades. Fortunately, the left boundary of the 'area to avoid' can be shifted towards the right, by increasing the LOA of the *information* stages while mostly delaying the automation of decision-making (Young and Stanton, 2023). As seen in Section 2.4.1, most automation issues arise when the LOA of the decision-making and/or execution stages are increased, leading to serial automation (Endsley, 2017). Automation should thus be designed to support rather than replace the human in the meantime, to ensure that human involvement does not drop below a certain minimum.

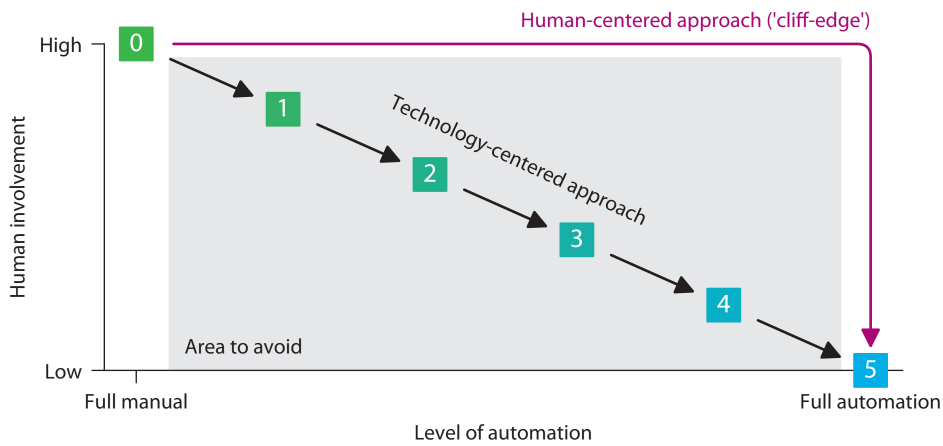


Figure 2.3: Human involvement in a human-centered 'cliff-edge' approach versus a technology-centered (i.e. function-based) approach, adapted from Young and Stanton (2023).

2.5 Constraint-based strategy

Another way to speed up the introduction of higher LOAs, loosely related to the cliff-edge principle (Young and Stanton, 2023), is to follow a constraint-based strategy. Although a relatively new phenomenon in the context of ATC, it is an established strategy in the automotive domain, with numerous companies developing fully autonomous self-driving cars that can operate in constrained environments (Kannan and Lasky, 2020). These companies skip the troublesome intermediate LOAs that lead to so many human-automation issues. When this strategy is applied to an ATC context two (or more) LOAs can be selectively assigned to a subset of the flights, as illustrated in Figure 2.1 by the two parallel but counterflowing vertical paths.

2.5.1 ATC example strategy: ARGOS

The ATC Real Groundbreaking Operational System (ARGOS) currently in development by MUAC is an example of a constraint-based automation strategy (Eurocontrol, 2024b). Where most ATC automation solutions proposed or implemented to date address individual tasks, such as conflict resolution, the ARGOS project takes a holistic approach, encompassing the entire set of ATCO tasks. This is reflected in the following three project objectives (Hendrickx and Tisza, 2019):

- Fully automated pre-tactical (planning) phase,
- Automated decision making and execution support for complex traffic scenarios,
- Fully automated control of basic traffic scenarios.

Here, the terminology 'basic' and 'non-basic' (or 'complex') flights is used to distinguish between flights requiring little ATCO attention and intervention (e.g., overflights), and flights requiring much ATCO attention, such as when they involve (multiple) vertical changes which may interact with multiple other flights.

The ARGOS philosophy is explicitly not to replace the ATCO completely, but to *"let ATCOs focus on the real, challenging work, to do what they are the best at, and leave the routine work to the machine"* (Hendrickx and Tisza, 2019, p. 1). Although this may seem very much like function-based allocation, there is a small but important nuance in the meaning of the word 'work'. Here it refers to the entirety of ATCO tasks for specific traffic scenarios, instead of specific (sub)tasks (e.g., only conflict detection) in any scenario. The workload reduction and other benefits that ARGOS is hypothesized to bring can then be used to work larger or busier sectors with the same number of staff (Lanzi et al., 2021).

MUAC envisions ARGOS to operate in one of three modes, pertaining to Levels 1–3 in Table 2.5 that can be enabled by the ATCO (supervisor). During the development of ARGOS, Level 1 will be progressively implemented to gradually gain ATCO trust and validate the systems, akin to a function-based strategy, before Levels 2 and 3 can be put into practice. Note that Levels 1 and 2 resemble several decision and execution LOAs from Table 2.4 to emphasize that the exact LOA in each stage can differ per situation and task. Figure 2.4 illustrates how the various ARGOS levels relate to each other on the LOA chart introduced in Section 2.3.1, with the vertical axis denoting the share of flights allocated at a certain LOA. Level 2 largely resembles a constraint-based strategy, where both ATCO and ARGOS fully control different parts of the traffic, and is the focus of this thesis.

Table 2.5: ARGOS and human ATCO responsibilities, adapted from Hendrickx (2023). Intermediate levels of the 10-level ARGOS taxonomy, that are not envisioned to be used, have been omitted for clarity.

			LOA per function (Table 2.4)	
Level	ARGOS	ATCO	Decision	Execution
0	N/A	Controls all flights.	0	0
1	Suggests plan for all flights. Executes approved plans for CPDLC flights. Reminds ATCO and defaults menu for non-CPDLC flights.	Approves, requests revision or rejects plans by ARGOS. Executes plans for non-CPDLC flights.	2-3	2-5
2	Presents and executes plan for flights allocated to it (by algorithm).	Controls all other flights. Takes back control over individual ARGOS flights.	2-3 & 6	2-5 & 7
3	Presents and executes plan for all flights. Alerts ATCO when flights are outside its comfort zone.	Monitor (stay in Level 3) when requested, or take back manual control for alerted flights (degrade to Level 2).	6	7
4	Controls all flights.	N/A	6	8

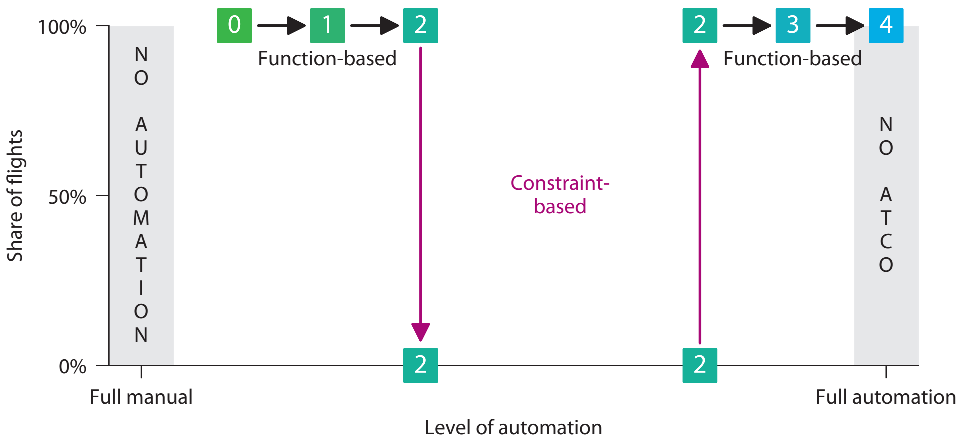


Figure 2.4: Level of automation chart for ARGOS.

A first step in introducing elements of ARGOS in the operational CWP was made in 2020 when occupied flight levels were color coded in the menus that ATCOs use to input their clearances, following the ecological interface design principle of intuitively displaying the constraints of the work domain (Borst et al., 2015). This feature was followed by the implementation of the Lateral Obstacle and Resolution Display (LORD) in 2024, a more advanced conflict space (constraints) display (Eurocontrol, 2024b). LORD shows the ATCO which combinations of altitude and heading clearances are free of conflicts in the coming 8 minutes and highlights any conflicting traffic. Both are examples of decision support tools that can be used in any traffic scenario.

2.5.2 Advantages and disadvantages

Compared to a function-based strategy, a constraint-based strategy circumvents the technology-centered stage where both human and automation are controlling the same flight at various stages. This cliff-edge approach (Young and Stanton, 2023) can speed up the development of fully automated systems, as many of the aforementioned human-automation challenges and their – often complex – required solutions are skipped.

Nevertheless, in practice, flights on either side of the LOA spectrum cannot be completely segregated, and thus inevitably create an overlap between agents, as introduced in Chapter 1. These interactions present a challenge that is also seen with self-driving cars, when they are mixed with conventional traffic (Nyholm and Smids, 2020) or pedestrians (Ezzati Amini et al., 2021). This calls for careful consideration on how to best facilitate the joint work. Determining which flights should operate at either the high or the low LOA is the main research area of this thesis.

2.6 Conclusions

When transitioning an ATC system towards higher LOAs, two main strategies can be taken. A constraint-based strategy, as seen elsewhere in the mobility domain, seems to be a promising way to avoid forthcoming issues when authority over flights is shared between human and automation. Unlike the traditional – and in ATC widely applied – function-based strategy, it enables attaining a high(er) LOA for a subset of flights. In doing so, it creates a parallel system where human and automation can work alongside each other. As true parallelism can only be approached, tailoring the overlap between either agent's work is essential to minimize interference in achieving a common goal. Determining which flights should be allocated to either human or automation is, therefore, of paramount importance.

As a first step, an empirical simulation experiment should be performed to explore the concept of sharing flights in a single airspace and see if ATCOs act alike in their preference for which flights should be allocated to which agent. Later experiments can then focus on finding and fine-tuning an algorithm for automatically allocating flights, such that this burden is not placed upon the ATCO.

3

Concept exploration

As discussed in Chapter 2, air traffic service providers around the world are aiming for higher levels of automation. So far, no extensive research has been performed on the feasibility and implications of selectively delegating specific flights to an automated agent, within one airspace. Through an exploratory experiment with six professional air traffic control officers (ATCOs), this chapter aims to provide some initial insights into the possibilities and complications of such a shared environment. Each ATCO was given suggestions from a distinct allocation scheme, but the system also allowed for manual revisions. Lessons learned in this chapter serve as a foundation for the remainder of the thesis.

An initial version of this chapter has been published in:

de Rooij, G., Borst, C., van Paassen, M.M., and Mulder, M. Flight Allocation in Shared Human-Automation En-Route Air Traffic Control. In *21st International Symposium on Aviation Psychology*, pp. 172–177. Online, 2021

An extended version, as presented here, has been published in:

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

3.1 Introduction

Air traffic control officers (ATCOs) work in a challenging and demanding environment. The continuous quest for more efficient and safer air travel drives the development of more advanced automation. Both Europe and the United States aim for higher levels of automation (LOAs) in the coming decades with a more supervisory/strategic role for humans (Prevot et al., 2012, SESAR Joint Undertaking, 2020). In such an environment, less people can handle more traffic in larger sectors. Despite striving for high levels of automation, humans are expected to play an important role in supervising these future systems and to intervene when automation falls short (Metzger and Parasuraman, 2005); people will ultimately remain responsible.

To be able to intervene, it is essential that ATCOs maintain vigilance, situation awareness and a sufficient skill level to perform tasks unassisted (Bainbridge, 1983). This could be achieved by not making the human a supervising bystander, but have them work side-by-side with automation in a team, both able to perform and share tasks. This sparks the question of what such co-operation should look like, and what impact it will have on human-automation performance.

Currently, the airspace is divided into sectors, each under the responsibility of a different ATCO. This requires considerable coordination between adjacent sectors and may lead to an imbalance in traffic load (and thus workload). To mitigate these issues, so-called flight-centric or sectorless operations are proposed (Birkmeier et al., 2016). Instead of coupling controllers to geographic areas, a single controller would be assigned to several flights, from departure to arrival, reducing the number of transfers and possibly providing a better workload balance. This, however, also introduces new challenges. Consider, for example, when two flights under control by different ATCOs are in conflict. Who should then solve the conflict?

Research suggests that conflicts are best solved at the planning stage (Hoc and Carlier, 2002), increasingly enabled by future initiatives like trajectory based operations (TBO, Enea and Porretta, 2012). However, even when flights are deconflicted before entering an airspace, unexpected conflicts may still occur, for example due to weather or emergencies (Corver and Grote, 2016), necessitating real-time ATC. In flight-centric ATC, it is expected that a computer algorithm then determines what flight is assigned to which controller, who then has to solve the conflict (Schmitt et al., 2011).

What if the other controller is not another human ATCO, but an automated computer system? How are flights then assigned to either agent: the controller or automation? Should problem-free flights be identified first (Drew and Makins, 1994) and subsequently automated? Should all flights involved in a conflict be controlled by either the ATCO *or* the automation, so as to mitigate additional workload related to inter-agent coordination? If not, who solves a mixed conflict? In addition, with an automated agent, it becomes possible to share (sub)tasks dynamically, back and forth, between human and automation. This could establish true teamwork, but only if the aforementioned questions have been answered first.

3

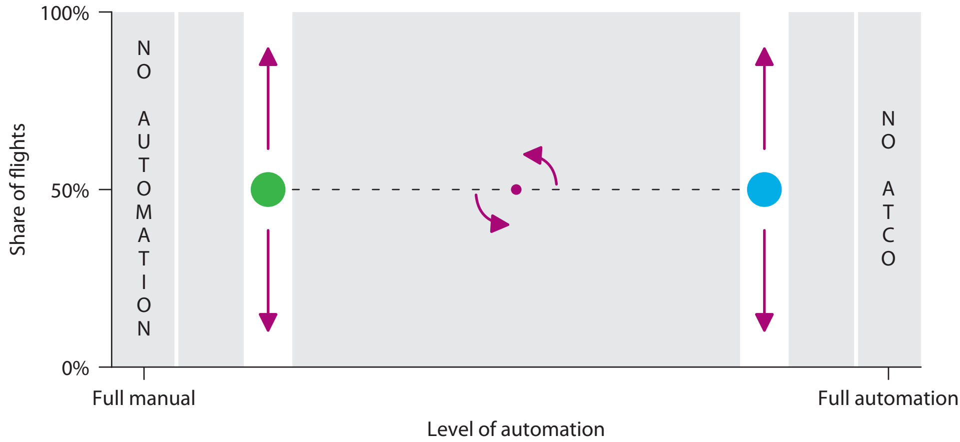


Figure 3.1: Experiment levels of automation, as introduced in Section 2.3. The distribution of flights between either LOA could be adjusted by the ATCO.

Existing research has primarily focused on allocating either the entire airspace to automation, or delegating certain tasks to an automated system (Martin et al., 2016). Research on delegating individual flights is scarce. Vanderhaegen et al. (1994) found that ATCOs do not act alike, when given the authority over which flights to automate. Some controllers allocated every other aircraft to automation, while others based the decision on their own workload.

As a first step, this chapter discusses an exploratory experiment with six professional ATCOs on the allocation of flights in a future shared human-automation en-route airspace. Initial system-produced suggestions were given for each flight, based on the notion of creating a parallel system (complete sectors allocated to either agent), or a more mixed system with flights in the entire airspace under mixed control. The ATCOs had the final say in which flights to delegate to automation, to gain insight in the conditions, such as sector-based structures, personal preference and automation capabilities, that could lead to a successful allocation strategy. Figure 3.1 illustrates how the ATCOs could only manipulate which (and how many) flights operate at either a low or high LOA, but that they cannot switch to other (intermediate) LOAs. That is, when the control point on the left moves down, the point on the right moves up and vice versa (i.e., around the pivot point in the center).

3.2 Method

3.2.1 Participants

Six professional en-route ATCOs (age M = 38.3, SD = 10.0, years of experience M = 14.8, SD = 8.7), from Maastricht Upper Area Control Centre (MUAC) voluntarily participated in a real-time simulator experiment. Four of them had active licenses for the Delta and Coastal (DECO) sector group, one for Hannover, and another for Brussels. During the initial briefing, they were informed about the content and aim of the study and were asked to sign an informed consent form. The experiment setup was approved by the Human Research Ethics Committee of TU Delft under number 1441.

3.2.2 **Apparatus**

SectorX, a TU Delft-built Java-based simulator (Figure 3.2) was designed to mimic the operational MUAC interface in order to ensure that participants could focus on working with the experimental automation (see Appendix A). Figure 3.3 shows the setup, consisting of a 1920 × 1920 pixels 27″ monitor and a standard computer mouse for control inputs.

Aircraft performance was modeled using Eurocontrol's Base of Aircraft Data (BADA) version 3.10 (Eurocontrol, 2012). The ATCOs could issue clearances for heading, route (direct-to or adding an intermediate waypoint) and altitude. Speed and vertical rates were not controllable. All clearances were uplinked through a datalink, removing the need for voice transmission over radio, and were instantaneously executed by the simu-

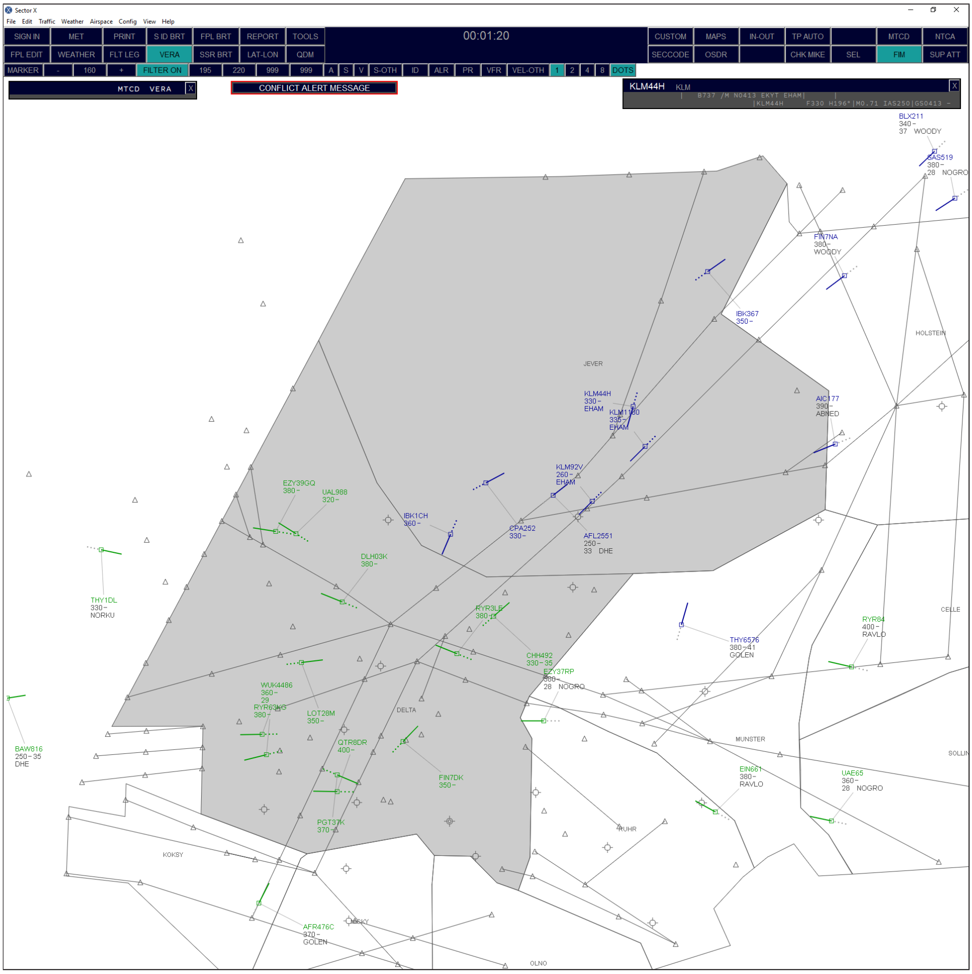


Figure 3.2: Simulator interface, with blue flights allocated to automation and green flights to the human ATCO. Note that flights approaching the controlled sector (gray polygon) have not yet been assumed, as indicated by their partially colored labels, which communicate a suggested allocation. Background colors have been inverted in the image for clarity.

3

Figure 3.3: Experiment setup.

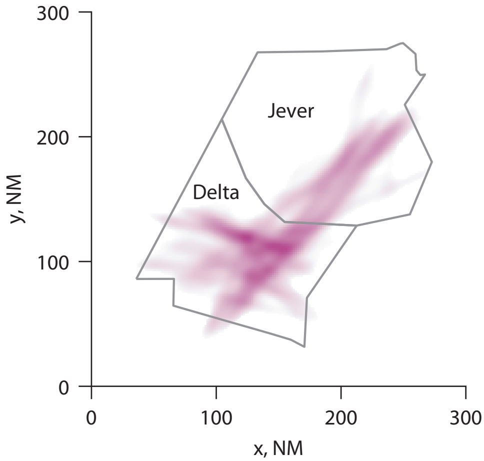


Figure 3.4: Traffic density of the scenario.

lated pilots. Using MUAC's verification and advice tool (VERA), the ATCOs could request the predicted minimum distance between any two flights (based on linear extrapolation of the current tracks) as well as the time-to-go until that minimum distance was reached. VERA's graphical representation of the conflict geometry and corresponding flight positions was not included in this particular experiment.

3.2.3 Airspace and traffic scenario

Participants were responsible for air traffic above FL245 in the combined Delta and Jever sectors above the Netherlands and part of Germany. This combination of sectors from the DECO sector group is not used in real operation, but was selected for the experiment because it encompassed two comparably sized sectors. Each ATCO experienced the same real-time traffic scenario: a typical radar snapshot of flights passing through the controlled sectors on an average day in February 2020, prior to the COVID-19 pandemic and the associated traffic reduction. Figure 3.4 shows the scenario's traffic density, with a clearly visible hotspot in the Delta sector.

There were between 15 and 30 ($M = 21$, $SD = 4$) flights in the airspace at any time (Figure 3.5), totaling to 104 flights for the entire 95-minute scenario. MUAC ATCOs would currently handle 20-25 flights in a combined sector the size of Delta and Jever. A higher peak value was chosen for the experiment to compensate for an expected workload reduction caused by offloading some flights to the automation and by the absence of voice communication. As two of the participating ATCOs were not licensed for this sector and the automation was new to all participants, the traffic density was not artificially increased beyond current levels (e.g., as sometimes done to simulate future traffic densities).

All flights followed standard routing or direct routes (24 flights) to their designated exit points and flew at a constant indicated airspeed of 250 knots. In addition to overflying traffic, arrivals and departures to several airports within or close to the sector were included. In total, 22 flights had to climb within the sector, 15 flights had to descend, and the remaining 67 flights had identical entry and exit levels. There was no wind.

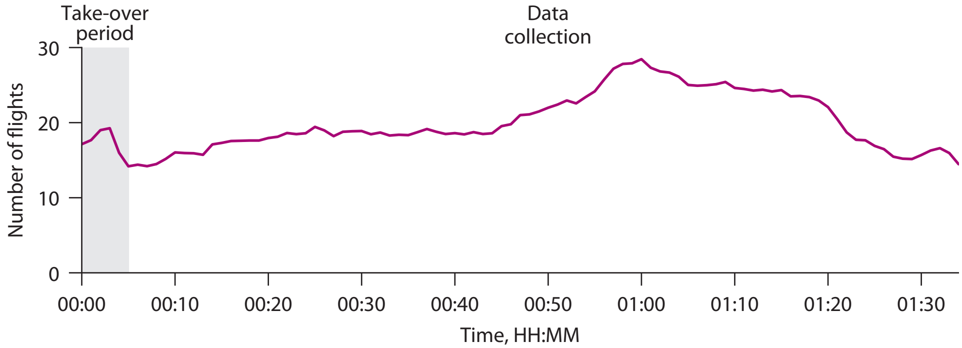
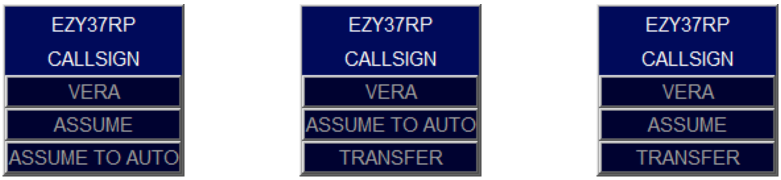


Figure 3.5: Time trace of the number of flights in the scenario.

3.2.4 **Automation**

During the exercise, the ATCOs were accompanied by an 'automated colleague'. When flights entered their sector, the ATCOs had to decide whether to manually assume the flight or delegate it to automation (Figure 3.6). The allocation remained flexible, allowing the participants to re-assume manual control or delegate flights at any time anywhere in the sector. This loosely corresponds to Level 1 or 2 of the ARGOS LOA model from Table 2.5, with the ATCO explicitly delegating flights to or taking back from the automation, but it lacks the proposed presentation of suggested actions or plans (i.e., there was no decision support for manual flights). To enforce an explicit transfer of responsibilities, all flights had to be manually transferred to the next sector, including those delegated to automation.



(a) Not yet assumed (b) Assumed manually (c) Assumed and delegated to automation

Figure 3.6: Callsign menu, shown when clicking the callsign in a flight label; ATCOs could delegate a flight to automation by pressing "ASSUME TO AUTO" or take it (back) manually by pressing "ASSUME".

For experimental control, simple, rule-based automated solvers were used to provide a basic level of automation that was both predictable and easy for the ATCO to understand. This diminishes or even removes the requirement for the highly independent automation to extensively communicate its intentions to the human operator, for which (complex) displays would be needed. The design of such displays is an entire research topic on its own, while this study's focus is the impact of flight allocation on teamwork rather than the impact of inter-agent communication.

The automation was fully capable of acting without human involvement and automatically executing actions to ensure safe air traffic. The automation performed the following tasks on flights delegated to it:

- Ensuring sufficient separation between automated flights (5 NM horizontally, 1,000 ft vertically).
- Delivering flights at their exit point and transfer level, climbing as early and descending as late as possible.
- Descending arrivals to FL260 for transfer to lower area control.

The automation solved conflicts between automated flights in the vertical plane only. It would never issue any heading clearances or direct-to's, meaning that flights allocated to automation would continue along their planned routes (or routes modified by the ATCO). If both conflicting flights were already at their planned exit level, one of the flights would be instructed to climb or descend 1,000 ft to solve the conflict. Thereafter, when clear of the conflicting traffic, the flight would be instructed to return to the exit level. All human-automation conflicts had to be solved by the ATCO, under the presumption that automation would not know the ATCO's plans or intentions. Apart from showing the issued clearances in the flight labels, the automation did not provide any other feedback on its intentions.

3.2.5 Procedure

All participants followed the procedure outlined in Figure 3.7, starting with a short briefing and pre-experiment questionnaire about their stance on automation. Next, each participant received a ten-minute training, during which the automation was introduced, they could familiarize themselves with the interface and practice their designated tasks. Both human-automation and automation-automation conflicts were shown to demonstrate how automation would handle both situations. The training was concluded with a short questionnaire.

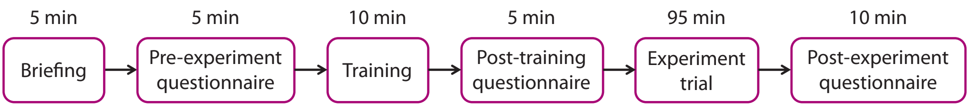


Figure 3.7: Experiment procedure.

Next, the experimental run started with a five-minute period simulating a shift takeover, during which scripted clearances were automatically executed without the ability to issue manual clearances, followed by 90 minutes of real-time simulation. Throughout the run, the experimenter observed and asked the ATCOs to explain their actions and what they were taking into consideration. Every three minutes, the ATCOs had to rate their instantaneous self-assessed (ISA) workload by clicking on an on-screen 0-100 scale (Tattersall and Foord, 1996). The scale showed their previous rating for reference. After the experiment, they completed a questionnaire with several open and Likert-type scale questions.

3.2.6 Independent variable

There was one independent variable, namely the suggested flight allocation, which was unique for each ATCO, as specified in Figure 3.8. The suggestions were either solely based on each flight's required absolute flight level change |ΔFL| from the entry level (NFL) until the transfer level (TFL) *or* based on the sector where the flight first entered the controlled airspace. Flights with |ΔFL| $\leq$ 2,000 ft were considered overflights in this context. For ATCOs 3 and 4, flights crossing the internal sector border would automatically change the responsible agent during the take-over period to adhere to the suggested allocation.

The suggestions were shown by the color of the flights' labels and radar position symbols upon approaching the sector (green = manual, blue = automated, inspired by the colors on the MUAC radar screen, where green flights are under control of the ATCO and blue flights are within that ATCO's area of responsibility but controlled by a different ATCO). The ATCOs were not briefed on which scheme was applied to them. They could ignore the suggestions and re-allocate flights at any time, even after delegating them to automation. Note that the chosen allocation color replaced that of the suggested allocation once a flight had been allocated by the ATCO.

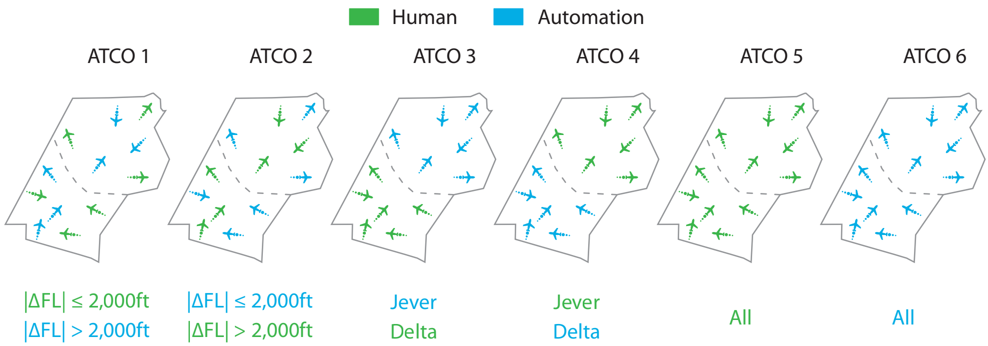


Figure 3.8: Suggested human-automation flight allocation schemes.

3.2.7 Control variables

The following control variables were the same for each participant:

- Airspace and traffic sample: as described in Section 3.2.3,
- Atmospheric conditions: international standard atmosphere without wind,
- Automation capabilities: as described in Section 3.2.4,
- No voice communication: all instructions were transmitted via controller-pilot data link communications (CPDLC) without transmission delay,
- No pilot delay,
- ATCO support systems: only VERA and short-term conflict alert (STCA).

3.2.8 Dependent measures

The following measures were collected in the experiment:

- *Pre-experiment questionnaire:* Prior to the simulation session, a short questionnaire using open and Likert-type scale questions aimed to probe ATCOs' stance and preconceptions on what automation could offer them. They were also asked to indicate which tasks and functions they would like or expect automation to support and/or take over.
- *Post-training questionnaire:* After brief exposure to automation in the new flight-based control allocation concept, the ATCOs expressed their initial opinions.
- *Experiment trial:*
  - Chosen flight allocations: Each ATCO was presented with a unique initial flight allocation, but was free to revise the suggested allocation as they saw fit by either delegating flights to automation or taking back control from automation at any time. As such, their level of appreciation for the initial allocation could be observed.
  - Control activity: The number, type and timing of issued clearances (altitude, heading and direct-to).
  - Perceived workload: Measured through an instantaneous self-assessed (ISA) rating (0–100) every 3 min during the trial (Tattersall and Foord, 1996),
- *Post-experiment questionnaire:* After the experiment, the ATCOs provided their opinions on the automation, flight-based control allocation concept, and simulation in general after having worked with it during the 90 min trials.

3.3 Results

Because the number of participants was small and the primary goal of the experiment was to provide a first insight into the feasibility and challenges of delegating part of the traffic to automation, we focused on providing a qualitative analysis of the raw data, observations and questionnaires rather than engaging in inferential statistics. Results are presented in accordance with the four data collection phases defined in Section 3.2.8.

3.3.1 Pre-experiment questionnaire

The results of the questionnaire at the start of the experiment revealed that the participating ATCOs had mixed opinions on automation and its involvement in their work (Figure 3.9). Answers into this and other questionnaire figures that are either relatable to human or automation have been colored accordingly: green corresponds to a more human-favorable answer, while blue relates to an automation-favorable answer. Most ATCOs trusted automation and expressed the opinion that it generally lowers their workload, but were also strongly of the opinion that a human should ultimately be in charge.

While this experiment focused on flight-based allocation, a human-automation team may also be created by sharing tasks according to a more conventional function-based allocation scheme. We replicated part of the study from Prevot et al. (2012) to see what kind of tasks the ATCOs would prefer to do themselves and which they would share with or completely delegate to automation if a function-based allocation was used. In line with that study, Figure 3.10 shows that the ATCOs indicated that a considerable number of tasks can be either shared with or completely delegated to automation. Transfer of control can be automated as a first step towards more automation, but ATCOs would like to be able to reject auto-transfers and to initiate early transfers. The ATCOs preferred to keep short-term, tactical actions under their control, while suggesting that more strategic long-term planning and routine tasks can be (partially) delegated to automation.

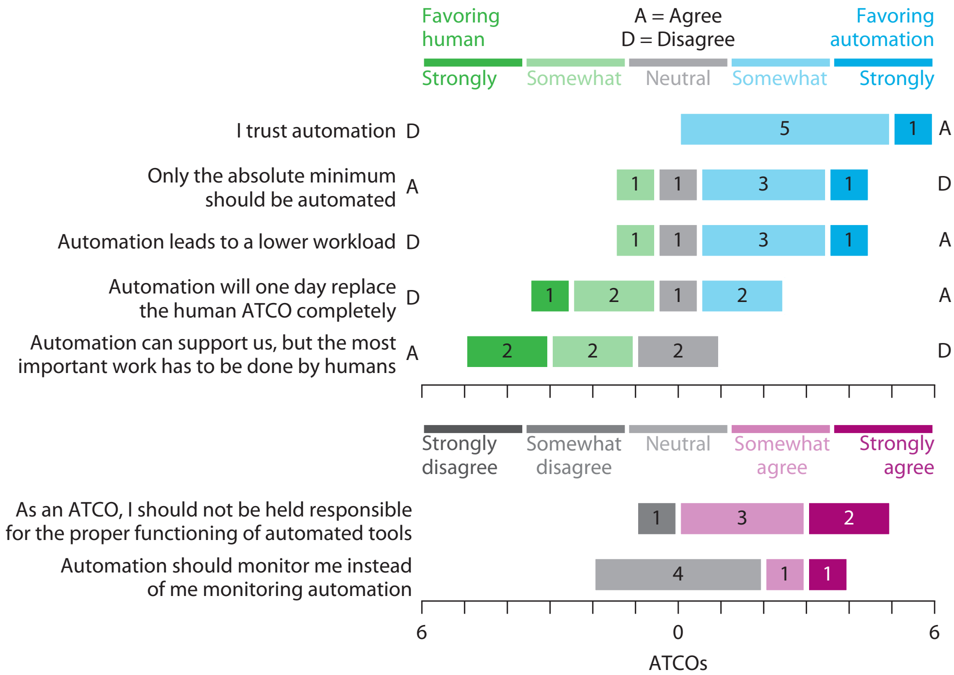


Figure 3.9: Pre-experiment ATCO response to various statements about automation in ATC.

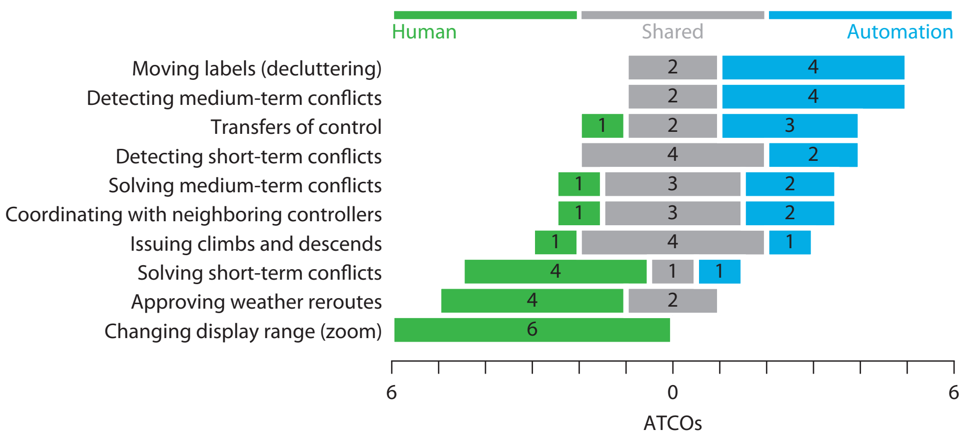


Figure 3.10: Human-automation task allocation as desired by the ATCOs in a function-based allocation system.

3

3.3.2 **Post-training questionnaire**

After a brief introduction and exposure to the experimental automation, the ATCOs had mixed opinions on whether it would be a useful asset in their work, as shown in Figure 3.11. Furthermore, contrary to their high level of trust in automation in general (Figure 3.9), the ATCOs indicated very low trust in this particular form of automation.

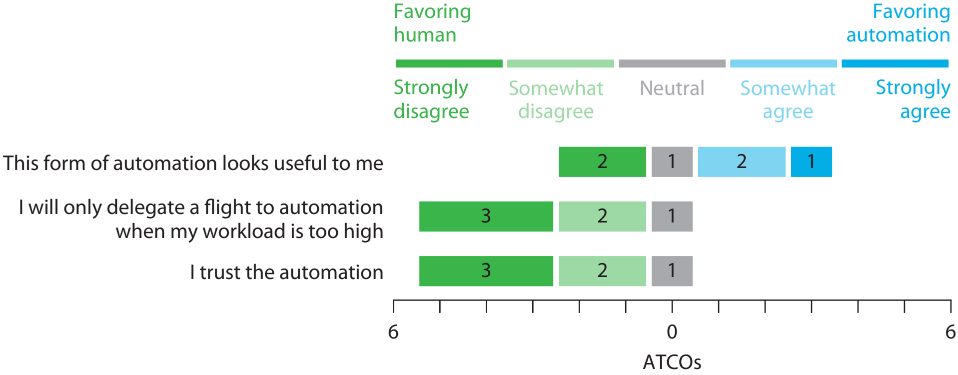


Figure 3.11: Post-training ATCO response to various statements about the experimental automation.

3.3.3 **Experiment trial**

Chosen flight allocation

All ATCOs delegated between 50% and 100% of the flights to automation, when considering the median share over the entire scenario (Figure 3.12). ATCO 3 appears to be an outlier, with a more balanced distribution compared to the other ATCOs, who were more automation-minded in their allocation decisions. For all ATCOs combined, flights were manually assumed for 23% of the total flight time.

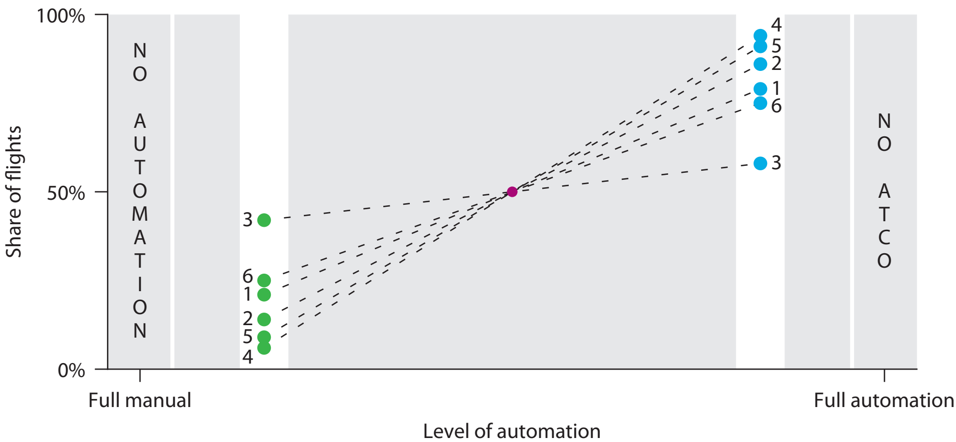


Figure 3.12: Allocation of flights to either LOA as chosen by the ATCOs. Numbers refer to specific ATCOs.

The evolution per ATCO over time, as shown in Figure 3.13, allows for a more detailed analysis. Had all ATCOs adhered to the suggested allocations, the figures on the right would have been the inverse of those on the left (i.e., all flights that were green for ATCO 1 would have been blue for ATCO 2 and vice versa). Note that despite identical traffic scenarios, the total number of simultaneously assumed flights varies slightly between ATCOs due to individual assume and transfer timings. Whereas most ATCOs largely ignored the suggested allocation, ATCO 3 tried to follow it after realizing that one of the sectors was completely handled by automation. Moving on to ATCO 4, the stark shift around 60 minutes from a nearly completely automated airspace towards a substantial number of manual flights followed a self-proclaimed *"test of the automation"* by this ATCO who deliberately re-directed flights manually. While the same ATCO claimed to be *"comfortable with purely monitoring a completely automated scenario"*, this 'test' may have been a sign of boredom. For ATCO 6, all flights were suggested for delegation; however, they did assume some of the flights (or at least part of their paths) manually. Note that none of the ATCOs who received the suggestion to delegate flights (everyone but ATCO 5) assumed more flights manually than was suggested to them.

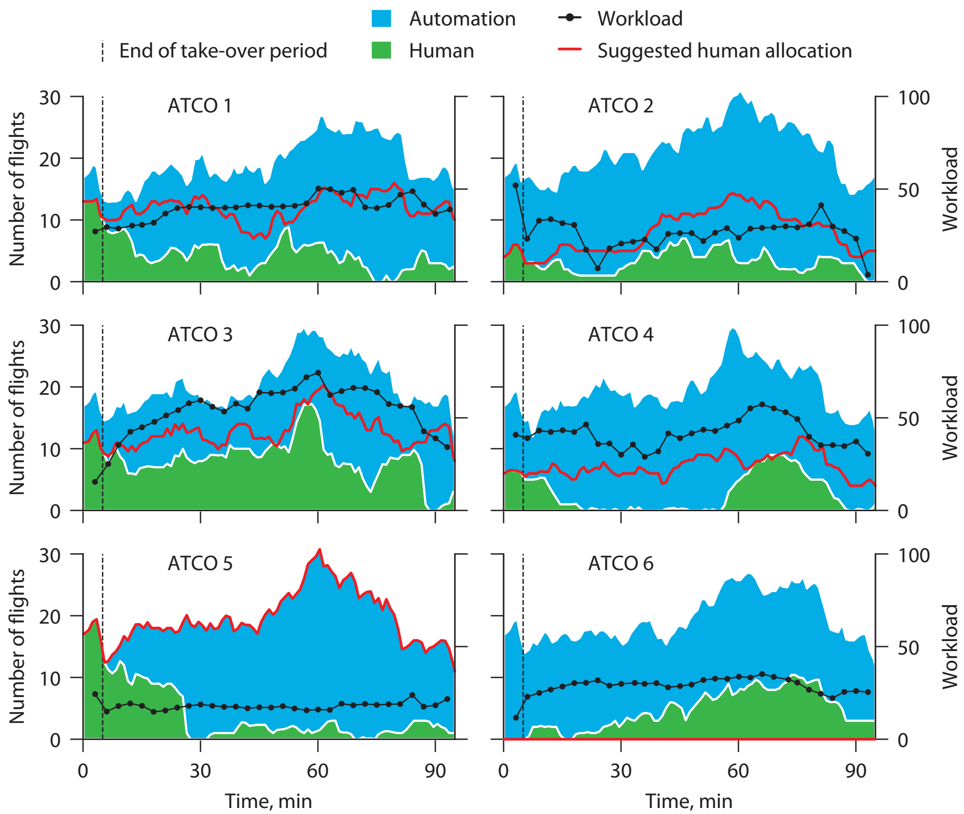


Figure 3.13: Stacked time traces of the number of actual and suggested flights allocated per agent and ISA workload ratings. The red lines correspond to the number of manual flights if the ATCOs had followed their uniquely assigned allocation suggestion from Figure 3.8.

The strategies suggested to the ATCOs and those they ultimately adopted can be more clearly seen when comparing the ground track density maps and required vertical movement $|\Delta FL|$ with the associated distribution of flight time (i.e., the duration that a flight was allocated to either agent) in Figure 3.14. Note that flights under control by either agent shared the same airspace in the experiment, and have only been split per agent in the figure for visualization purposes. The left (green) and right (blue) figures always add up to 100% of the traffic for each ATCO. As an example, ATCO 1 delegated 73% of the flight time in the Jever sector and 81% of the flight time in the Delta sector to automation. For the remaining time, 27% and 19% respectively, the flights were under manual control. In terms of vertical movement, only 24% of the flight time of flights with $|\Delta FL| \leq 2{,}000$ ft (LT) was under manual control by this ATCO, with the remaining 76% being automated. Following the suggested allocation would have led to 100% manual flight time, as indicated by the green outline of the 'LT' bar. The 'GT' bars resemble the flights with $|\Delta FL| > 2{,}000$ ft. The suggestion was that ATCO 1 automate all of these flights, they instead chose to take manual control for 17% of the flight time.

As already hinted at by Figure 3.13, ATCO 3 exhibited the greatest adherence to the proposed allocation strategy among all ATCOs. This ATCO even delegated flights as they transitioned from Delta *to* Jever, commenting that solitary manual flights in a predominantly automated area were difficult to handle. This resulted in 95% of the flight time in Jever being delegated to automation, approaching the suggested 100% in this sector-based allocation. Interestingly, the same ATCO did not consistently re-assume automated flights that entered Delta *from* Jever, resulting in a considerable 46% of the flight time in Delta being delegated to automation instead of the suggested 0%. All other ATCOs had a considerably more even distribution in both sectors.

For ATCOs 1 and 2, the suggested allocation was based on the required $|\Delta FL|$ between sector entry (NFL) and transfer (TFL) rather than on the entry sector. ATCO 2 appears to have followed the suggestions slightly better than ATCO 1, as shown by a 97% delegation of overflights and considerably lower share of 66% for flights with $|\Delta FL| > 2{,}000$ ft. It must be noted that the bars in Figure 3.14 are based on the $|\Delta FL|$ between sector entry and exit; therefore, flights that were flying level for the majority of their flight time may have predominantly contributed to the blue GT bar even though they were assumed manually during the actual (short) climb or descent phase. Upon closer inspection, all ATCOs frequently delegated flights as soon as they had (almost) reached their TFL and were clear of any remaining conflicts. ATCO 3 is again a noticeable outlier, with manual control used for a relatively large share of flights with small $|\Delta FL|$ for a prolonged time (43%) due to this ATCO's adherence to the suggested *sector*-based allocation.

One of the ATCOs kept a low and slow flying Beluga cargo flight with $|\Delta FL| = 0$ ft under manual control throughout the sector. This ATCO commented that this was mainly because of the flight's close proximity to lower airspace that is not controlled by MUAC and the potential disruption it may cause to or receive from traffic in that airspace. If such traffic had been present in the experiment, the other ATCOs may have also decided to manually control the Beluga.

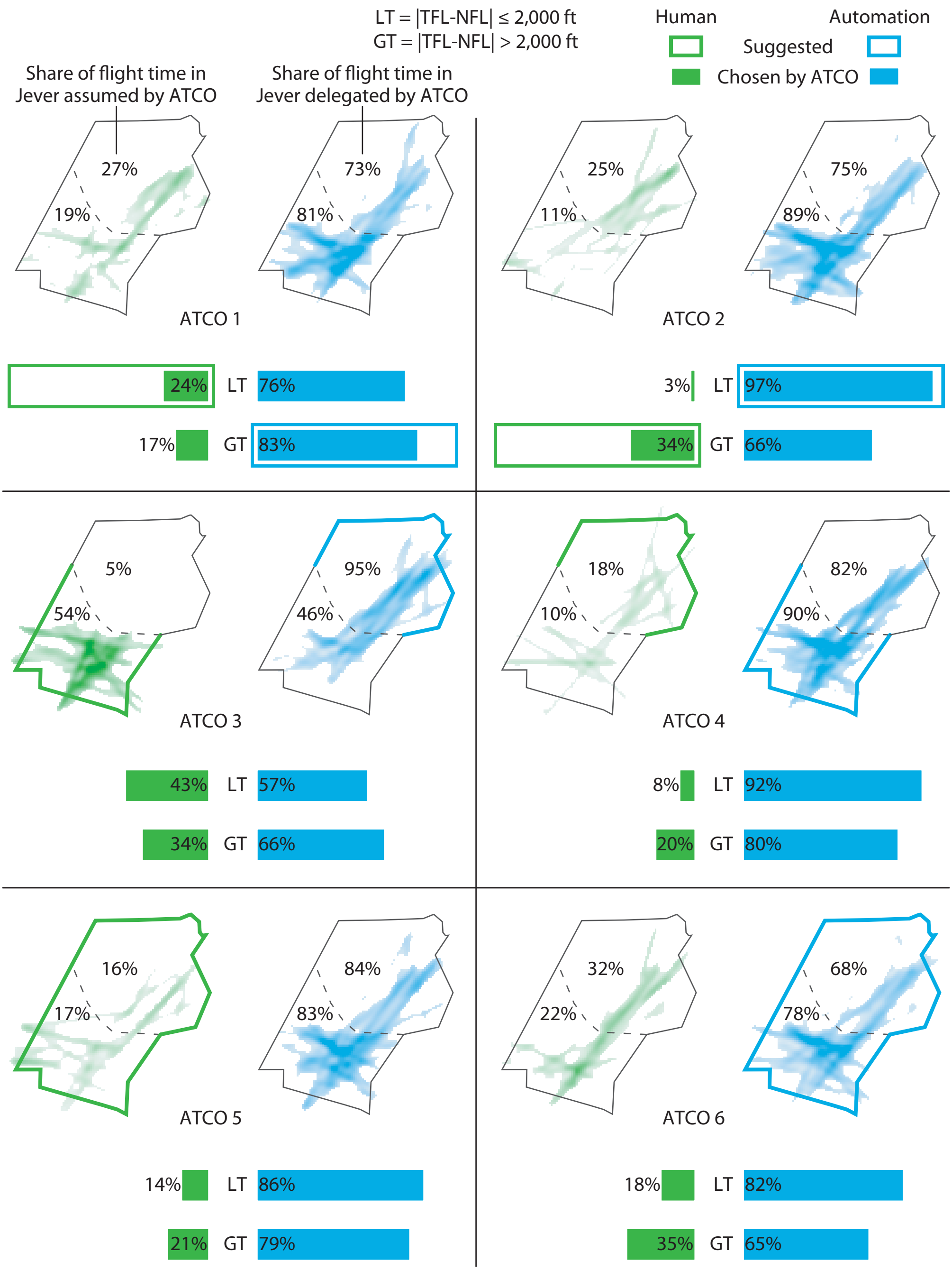


Figure 3.14: Flight density maps and share of flight time as allocated by the ATCOs, split per agent. The colored outlines indicate the suggested allocation from Figure 3.8, i.e., the absolute required flight level change |ΔFL| between sector entry (NFL) and transfer to the next sector (TFL) for ATCOs 1 and 2, entry sector for ATCOs 3 and 4, and all manual or automated for ATCOs 5 and 6.

Control activity

The ATCOs issued 51% of all clearances (30% of altitude clearances), leaving the rest to the automation (Figure 3.15). As discussed in Section 3.3.3, most ATCOs took manual control for a short period of time to issue a clearance before delegating the flight for the remainder of its trajectory. Here, 55% of the flights did not change agent after being assumed, while 43% of the flights that did spent less than one minute with the ATCO (Figure 3.16). This was especially true for flights that could benefit from a direct-to, which the automation could not issue. Interestingly, ATCO 3 hardly sent any flights on a direct-to, while ATCO 5 did so for over 50 flights. ATCOs 1 and 3 both issued intermediate-level clearances to up to 25 flights, resulting in an above average total number of altitude clearances.

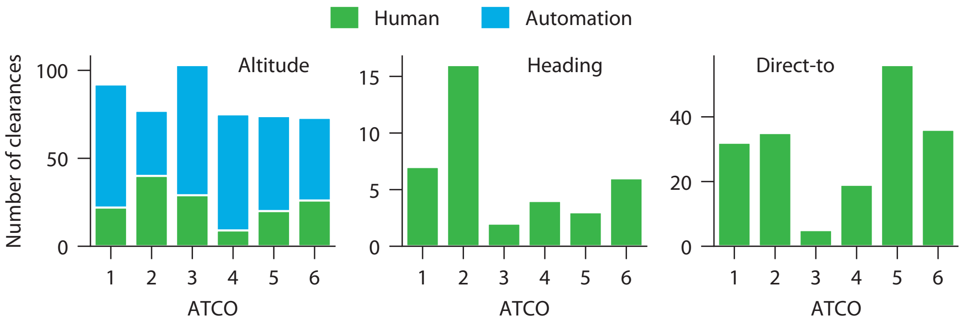


Figure 3.15: Number of issued clearances per ATCO and agent.

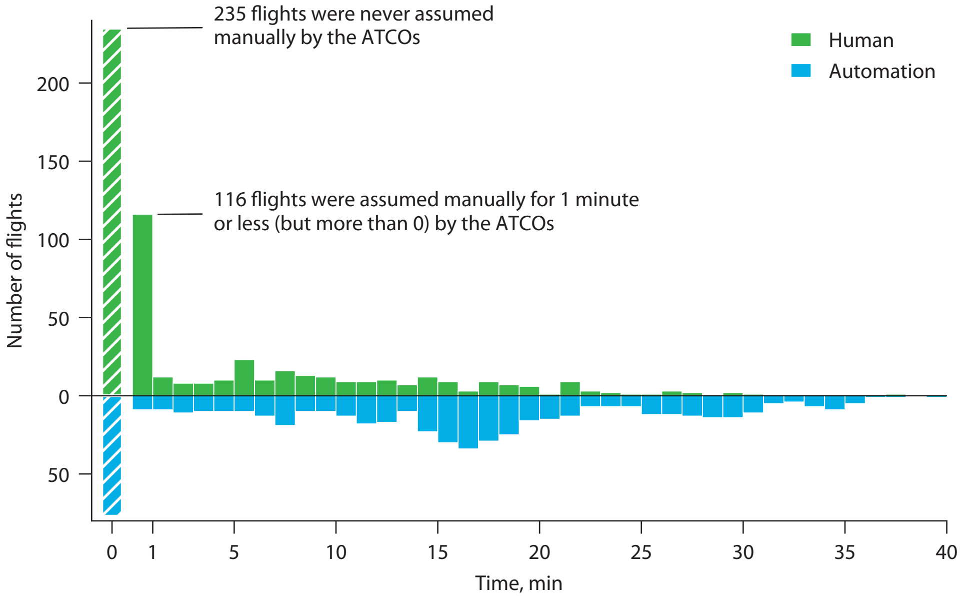


Figure 3.16: Histogram of the total duration that flights were either under manual or automation control. All flights are included twice per ATCO, once for either agent.

Perceived workload

The ISA workload ratings in Figure 3.13 do not show a significant correlation with the fraction of automated flights. A higher share of automated flights did not provide the workload reduction that ATCOs generally expect with automation (Figure 3.9). ATCO 5 reported a very consistent workload, ranging between 15 and 24 (on a 100-points scale), whereas for ATCOs 3 and 4 it varied considerably. In general, the ATCOs commented that their workload was low due to a relatively low traffic demand (when considering that part of the tasks were performed by automation) and the absence of voice communication. Because every ATCO only experienced one of the suggested allocation schemes, and as workload ratings are subjective, no further within-participant comparisons can be made with respect to the suggested or followed allocation schemes.

Despite 51% of the control actions being performed by the ATCOs, only 23% of all flight time was manually assumed. This discrepancy may explain why the number of flights allocated to either agent did not correlate with the perceived workload.

3.3.4 **Post-experiment questionnaire**

Flight allocation

At the end of the experiment, the ATCOs were asked what percentage of flight time they believed themselves to have delegated to automation over the entire run. All ATCOs were able to estimate this within eight percent point of the actual median (Table 3.1), indicating a good match.

Table 3.1: Self-reported and actual flight time delegated to automation over the entire run.

	ATCO						
	1	2	3	4	5	6	Mean
Self-reported (%)	71	86	66	92	95	71	80
Actual mean (%)	77 (+6)	85 (-1)	60 (-6)	88 (-4)	80 (-15)	77 (+6)	78 (-2)
Actual median (%)	79 (+8)	86 (=)	58 (-8)	94 (+2)	91 (-4)	75 (+4)	81 (+1)

The questionnaire provided further insight into how ATCOs determined whether flights should be delegated in the experiment trial (Figure 3.17). Note that the ATCOs could classify factors as ‘not considered’ (this ranged from three to six per ATCO), meaning that not all ATCOs ranked the same number of factors. Traffic directly surrounding a flight was considered especially important when there were many nearby manual flights. Delegating a single flight to automation would then have added (too) much uncertainty. Overflights are generally considered to be more predictable than arrivals and departures, making the type of flight another important factor. Special flight types, such as the slow and low-flying Beluga cargo flight, also played a role here. The suggested allocation was given low priority or even ignored by most ATCOs. ATCO 3 ranked this as the most important factor and acted accordingly, as confirmed by Figures 3.13 and 3.14. If automation would have been capable of giving direct-to’s, the ATCOs commented that they would have delegated more flights in this experiment. Four ATCOs included the automation capabilities, while only two included their workload (which was relatively low, as discussed

3

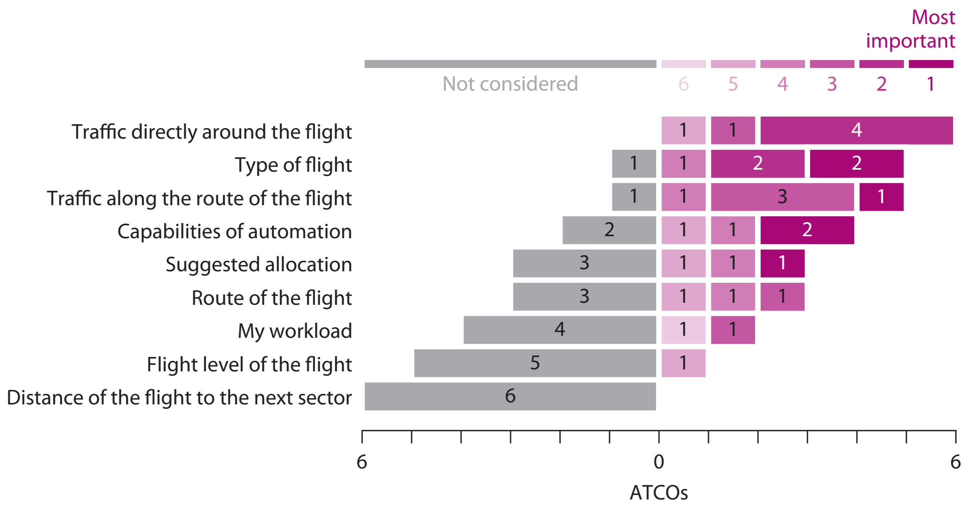


Figure 3.17: Factors driving the ATCOs' decision-making on whether to allocate flights to automation or to themselves.

in Section 3.3.3). Surprisingly, some ATCOs mentioned that the distance to the next sector should be included in an allocation algorithm, despite none of them considering it in their own allocation decision-making during the experiment.

Although every ATCO only received suggestions from a single flight allocation scheme in the experiment trial, the questionnaire asked their opinion on all of the schemes from Figure 3.8. Grouped per type of allocation-driver, their feedback was as follows:

- **Vertical change:** the ATCOs unanimously agreed that 'complex' climbing/descending flights need to be handled manually (potentially with support tools). They indicated a strong preference for delegating 'basic' (over)flights to automation. For most ATCOs, this was also reflected in the time that they delegated such flights to automation: 70% of the flights with identical entry and exit levels ($|\Delta FL| = 0$) were automated for 95% of their flight duration, versus only 50% of the flights requiring some level change. Although some ATCOs commented that a 5,000 ft level change would have been a more appropriate threshold to divide traffic in basic and complex than the used 2,000 ft, this was not directly reflected in their chosen allocation strategy. All traffic that had to change levels evoked more manual control than overflights, and as such could be considered at least somewhat 'complex'. This corresponds to the allocation suggested to ATCO 2.

- **Sector:** allocating flights per sector was outright rejected by three ATCOs, who commented that the choice of whether or not to delegate a flight should depend on the situation rather than the geographic sector. Two ATCOs (including ATCO 3) did see some use in it when one of the sectors was busy and/or required more concentration, while one ATCO refrained from commenting.

- **Full manual or automation:** four ATCOs praised the fully manual scheme for giving them full authority over which flights to delegate to automation and when (e.g., after turning and climbing). One ATCO preferred to have overflights always proposed to automation, while the remaining ATCO simply disliked this scheme. Finally, the fully automated scheme received favorable comments from five ATCOs, provided that the automation functioned well and that the supervising ATCO could take over at any moment. One ATCO criticized it on the basis that there will always be flights that need human involvement due to their flight profile or because they pass through traffic hotspots.

Perceived impact of the automated agent

Figure 3.18 shows that the ATCOs believed the automation as implemented in the experiment to have somewhat worsened their situation awareness and work style. Nevertheless, all ATCOs classified their situation awareness as 'okay', the middle score on a five-point Likert scale from 'poor' to 'very good'. All ATCOs mentioned that they paid (much)

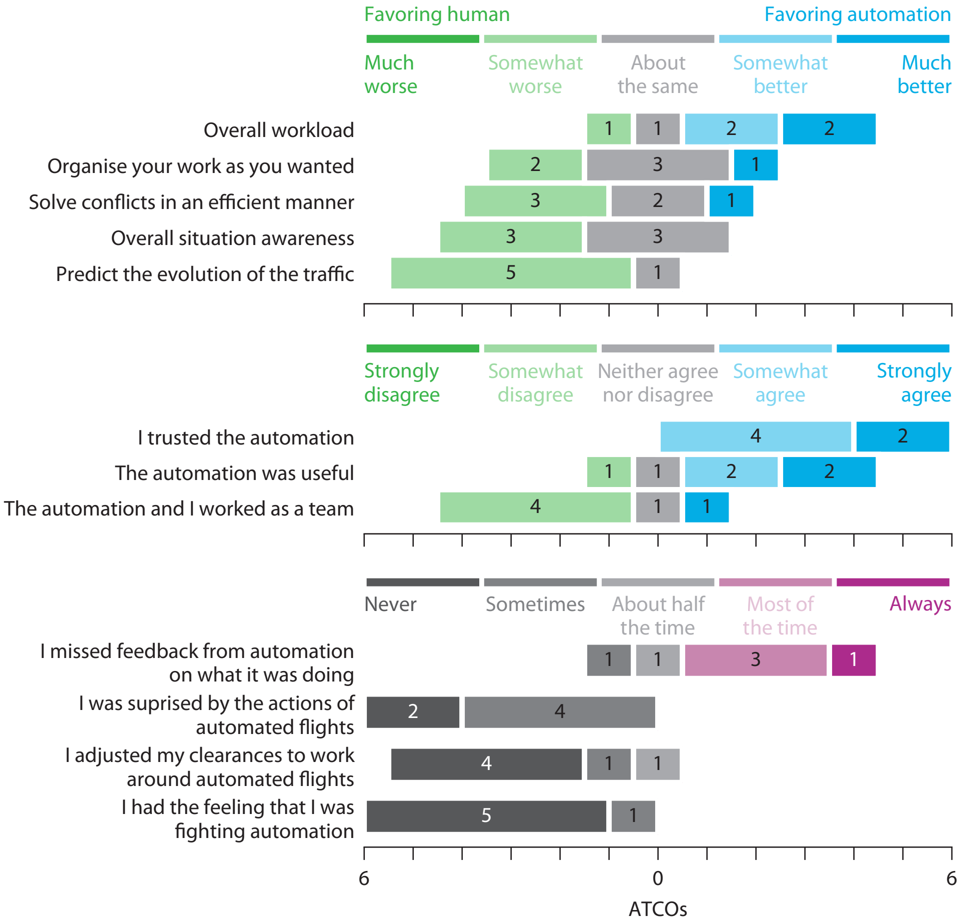


Figure 3.18: Post-experiment ATCO response to various statements about the impact of automation.

less attention to the blue automated flight, akin to transferred flights, even though they knew that they were still responsible for these flights. At the only (not explicitly programmed) occurrence of a short-term mixed conflict in the experiment, the involved ATCO was surprised by the STCA. He explained that he had not spotted the conflict because the automated flight was emerging from *"a sea of blue aircraft"*. While the ATCOs were aware that such conflicts could occur and would require their involvement, the unexpected occurrence could be considered a 'gray swan' as described by Wickens (2009). Any other potential mixed conflicts were avoided by the ATCOs in a timely way.

The ATCOs reported the lack of feedback from the automation regarding what actions it would take and the associated uncertainty as the largest contributor to their negative experience. All ATCOs would have liked the automation to at least show its intentions about where on the trajectory it would start and end a climb or descent. Interestingly, they also reported that the automated agent's actions were not surprising, suggesting that the implemented rule-based automation was in fact predictable. However, several ATCOs reported that they proactively took manual control at times (e.g., when approaching a top of descent) in order to forestall any potential surprises from automation.

In the end, the automation did have a positive impact on their overall perceived workload and was considered to be somewhat useful by most ATCOs, exceeding expectations as reported in the post-training questionnaire (Figure 3.11). Nonetheless, only ATCO 4 considered collaborating with the automation to be a form of teamwork.

Simulation fidelity

This experiment was a first test of the SectorX simulator's MUAC style. The ATCOs were asked to rate the fidelity of various aspects when compared to the operational human-machine interface and traffic (Figure 3.19). The most missed interface feature was VERA not showing the conflict geometry at the closest point of approach[1].

The aircraft behavior was rated as unrealistic by two ATCOs, who particularly mentioned the simulated climb rates. This can be partly attributed to the fact that in the absence of operational data, all flights were assigned a constant reference mass from the BADA performance model, making them exceptionally light on departure and heavy on arrival. All aircraft were flying at a constant indicated airspeed, and did not follow the standard climb/descent profiles. In addition, no pilot delay was modeled, meaning that all clearances were immediately executed.

Finally, the traffic scenario was considered realistic, although somewhat low in traffic, leading to relatively low workload ratings throughout the runs.

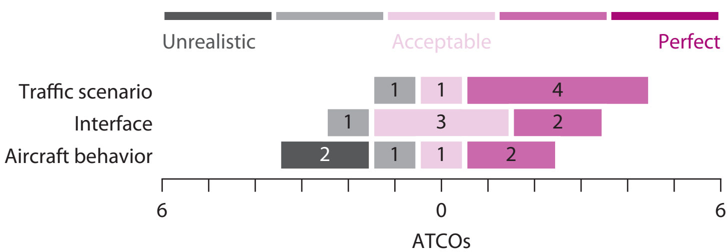


Figure 3.19: Post-experiment simulator fidelity ratings.

[1]This feature has since been added to the simulator and was available in the other experiments of this thesis.

3.4 Discussion and recommendations

The experimental results of this initial exploration of flight-based control allocation in en route ATC show promising results in terms of concept feasibility and ATCO acceptance. Followup studies should address the limitations and assumptions of our study, e.g., by introducing pilot delays, wind, and voice communication. Together with a more demanding traffic scenario, in which unaided manual ATC would lead to excessive workload, this is hypothesized to better demonstrate the benefits of offloading flights to automation in conjunction with the distinctive problem-solving abilities of ATCOs. In addition, it is recommended that future research focuses on the two research areas outlined below to bring the concept one step closer to operational implementation.

3.4.1 Automation

At the start of the experiment, all ATCOs reported having a high level of trust in automation in general, but were nevertheless suspicious of the experimental automation after the (short) training. Despite this, their trust was largely restored through the experiment trial (Figure 3.20). According to the ATCOs, this trust buildup was mainly due to their seeing the automation perform well. The rule-based form of automation (programmed to be 'perfect'), clear separation of responsibilities, and absence of uncertainties such as wind and pilot behavior further contributed to this.

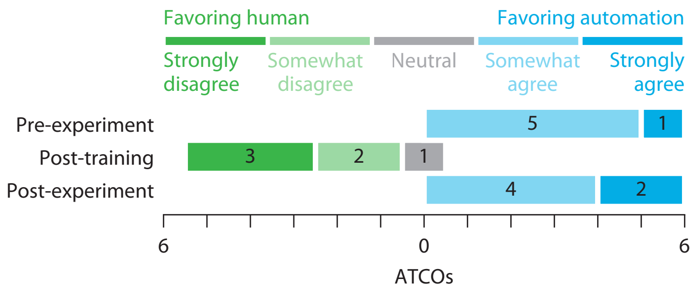


Figure 3.20: The ATCOs' stances on "I trust (the) automation" at three different moments during the experiment.

Fostering a high level of trust in the automation is of paramount importance in successfully creating a parallel system. In this experiment, the ATCOs needed some time to observe and monitor the behavior of their digital colleague before a sufficient level of trust was gained. Occasionally, they saw a need to intervene by taking back control over a flight, issue a clearance, and delegate it back to automation again. Monitoring and intervention are typical of supervisory control environments, which invoke the more serialized interactions that flight-based control allocation seeks to avoid.

Based on ATCO comments and observations during the experiment, the need for monitoring and intervention appeared to be caused by the limited capabilities of automation combined with the ATCO's responsibility for resolving mixed conflicts as well as the rather simple and pragmatic flight allocation schemes. To achieve a more desirable parallel system, automation should be able to perform all ATCO tasks and should have the responsibility and authority to resolve mixed conflicts. The potential downside of the latter, where the automation avoids all other automation- and human-directed flights, is that the two agents might engage in oscillating behavior by reacting to each other's actions. One way to avoid this issue is to use a smarter flight allocation strategy based on predicted interactions between flights rather than on individual flight and/or sector characteristics. If mixed conflicts do occur, a 'less impacted flight algorithm', as developed for flight-centric ATC (FCA, Finck et al., 2023b) where flights stay with a single ATCO throughout their trajectory, could quickly identify which agent can resolve the conflict with minimal effort and disruption and alert the ATCO if it is their responsibility.

3.4.2 Flight allocation suggestions

In general, the ATCOs did not follow the suggested flight allocations, with half of the ATCOs explicitly reporting that they ignored them (Figure 3.17). The suggestions were based on simple pragmatic schemes, and did not take into account the actual interactions between flights. While the ATCOs indicated that the automation capabilities and suggested allocation were equally important, the former seemed to be prevalent in their chosen allocations. The high number of lateral control actions issued by the ATCOs illustrates the lack of lateral control ability on the part of the automation. Nevertheless, half of the ATCOs reported that they did not consider the automation's capabilities in their allocation.

To provide more fitting (and consequently more accepted) flight allocation suggestions, it may be beneficial to take the actual (predicted) complexity of flights into account. In contrast to the abundance of research on sector-based complexity (e.g., dynamic density, interval complexity, fractal dimension, input/output approach, Lyapunov exponents and trajectory-based complexity (Prandini et al., 2011, Prevot and Lee, 2011)), the complexity of individual flights is less understood. Flight-centric complexity is a prerequisite for the automated allocation of flights in both FCA and our proposed operations. In FCA, a predicted workload increment per flight is used to allocate flights while balancing the workload between ATCOs (Finck et al., 2024), whereas in our proposal low-complexity flights are allocated to the automated agent. In both cases, interacting flights are best assigned to the same agent in order to prevent 'mixed conflicts' between flights under the responsibility of different agents and avoid excessive communication and coordination efforts between agents.

The complexity of a flight is not necessarily constant throughout its time in the sector. Section 3.3 showed that ATCOs frequently delegated flights after they had passed the 'challenging' part of their route, i.e., the climb, descent or conflict situation. External factors such as adverse weather conditions and associated reroutes can also play a role as trajectory uncertainty increases. Thus, flights could be re-allocated when their complexity changes beyond a certain threshold.

3.5 Conclusions

This exploratory study has yielded useful insights into human-automation teaming in a realistic ATC setting. We show that, after initial skepticism, professional en-route ATCOs are not averse to sharing their work in an airspace with automation. In a simplified situation lacking uncertainties due to wind, emergencies and pilot requests, a high level of delegation to automation was reached under the condition that flights were on direct routes and free of conflicts. The ATCOs generally ignored the suggested allocations, hinting not only at the need for a different form of allocating flights but also at the impact of automation capabilities.

While some ATCOs may simply prefer to make their own division when flights come in, the majority of the participants welcomed an automated allocator. However, the automation should be able to perform all tasks in order to prevent the serialized interactions found in systems requiring human supervision. This includes issuing direct-to's, assuming/transferring flights and solving or communicating about conflicts between human-directed and automation-directed flights. To further minimize interactions between human and automation, future research should take a closer look at determining the complexity of individual flights and consequently classifying them as 'basic' or 'complex' such that a fitting allocation scheme can be applied. Together with empirical studies on the various forms of task sharing and distribution, this can help establish human-automation teamwork in a shared ATC environment.

4

Empirical task analysis

While it is not expected that air traffic control officers (ATCOs) will completely ignore flights delegated to the automation, findings from Chapter 3 and previous research suggest that they do pay less attention to those flights. Based on literature research and workplace observations, this chapter first presents an overview of the cognitive workflows of ATCOs in the conflict detection and resolution task. To support this with further empirical insights into the impact of flights that lack a proper representation in the ATCO's mental model, a worst-case scenario is simulated in which the ATCO has been completely out-of-the-loop with respect to the automation-directed flights.

The first part of this chapter was published in:

de Rooij, G., Tisza, A.B., Borst, C., van Paassen, M.M., and Mulder, M. Towards Human-Automation Teamwork in Shared En-Route Air Traffic Control: Task Analysis. In *IEEE International Conference on Human-Machine Systems*. 2022. ISBN 9781665452380. doi:10.1109/ICHMS56717.2022.9980715

It has been supplemented here with an empirical quantification of the presented task workflow analysis.

4.1 Introduction

The air traffic management (ATM) community is increasingly advocating higher levels of automation (LOAs) to improve efficiency and capacity, with a more strategic supervisory role for humans (Prevot et al., 2012). The Single European Sky ATM Research (SESAR) program envisions a future in which air traffic control (ATC) tasks are increasingly automated, starting with information acquisition and analysis, followed by action implementation, and finally decision making (SESAR Joint Undertaking, 2024). Concurrently, the associated systems should become more autonomous, initiating a greater number of actions without human intervention. Nevertheless, air traffic control officers (ATCOs) are expected to play an important role in supervising these systems and to intervene when automation falls short (Metzger and Parasuraman, 2005).

Decades of human-automation research have taught us that bluntly shifting tasks previously carried out by humans to automated systems is not the way forward (Endsley, 2017). The manual execution of certain tasks can actually be beneficial in other tasks. Delegating only the conflict detection task to automation, for example, while leaving the resolution task for the ATCO, leads to a situation awareness (SA) reduction and increases the use of less optimized resolutions (Mercer et al., 2017). Keeping ATCOs or any other operator actively engaged is key in preventing many of the issues encountered when introducing higher LOA, such as skill degradation and reduced SA (Strauch, 2018). In the past decade this insight has led to a growing interest for, and belief in human-automation *team*work, with human operators dynamically sharing tasks with automation (O'Neill et al., 2022).

Where considerable research is devoted to (dynamically) allocating certain *functions* to automated ATC systems, Eurocontrol's Maastricht Upper Area Control Centre (MUAC) takes interest in a slightly different approach (Hendrickx and Tisza, 2019). As a first step towards higher LOA, *part of the traffic* may be completely directed by an automated system to alleviate the workload of ATCOs and increase capacity or sector size. A prime candidate for such delegation are flights that can be considered 'basic' (i.e., requiring little monitoring or cognitive effort). When these basic flights are delegated, the ATCOs can focus on 'non-basic' flights that require more attention and skills which humans are known to be good at (Endsley, 2017). Putting problem-free flights that do not require any action at all in a separate group was already proposed in the 1990s (Drew and Makins, 1994). Its associated workload-relief was limited though, as ATCOs already pay relatively little attention to these flights. Therefore, it is desirable to delegate flights that *do* require some active control. An example of more recent research focused on US-based operations and mixed self-separating flights with human-directed flights (Prevot et al., 2005), requiring considerable airborne equipment and wide adoption of time based operations.

The delegation of some flights in a sector introduces a number of challenges. For example, the question of which agent is responsible for solving a conflict when the involved flights are directed by different agents, a mixed conflict. Research on flight-centric operations (also known as sectorless operation), where ATCOs are responsible for flights from start to finish rather than based on geographical sectors, seems to allocate mixed conflicts to either ATCO, based on pre-determined (conflict) criteria (Korn et al., 2020) or ATCO workload (Schmitt et al., 2011). When one of the flights is instead directed by an automated system under the supervision of an ATCO who controls the other flight, a different approach may be more beneficial. The ATCO may, for example, prefer to manually solve the conflict and prevent automation from working against their plan for dealing with flights in their sector. ATCOs not actively involved with a considerable share of traffic in their sector may experience a detrimental effect on their SA.

In previous work (Chapter 3), a preliminary setup was experimentally tested where ATCOs could delegate individual flights to an automated system. While it showed the feasibility of such a shared airspace and its acceptance among ATCOs, it also revealed that ATCOs adopted different allocation strategies than anticipated. Their seemingly reduced attention for delegated flights also suggests these are not well present in their mental models, complicating conflict detection and resolution (CD&R) of mixed flight pairs.

To shape the implementation of a shared human-automation airspace, it is paramount to better understand the implication of delegating (part of) the traffic to automated systems. This requires a thorough understanding of the tasks carried out by ATCOs. In 1999, Eurocontrol published an integrated task analysis (ITA), based on interviews, observations and flight progress data obtained at five en-route control centers in Europe (Dittmann et al., 2000, Kallus et al., 1999). While it provides an extensive insight into the generic tasks of an en-route ATCO, it lacks on several aspects that would be useful here for shaping future human-automation teaming. For example, how the task flows change in the presence of automation and consequential mixed conflicts, as introduced in the concept of operations. In a similar way, the ITA lacks how current-day support tools are increasingly utilized and where they fit in the processes. Finally, it lacks temporal quantification, making it difficult to objectively assess the performance and workload impact of different allocation strategies.

After first presenting a concept of operations (CONOPS, Section 4.2), this chapter uses the Eurocontrol ITA as an inspiration to introduce flowcharts in Section 4.3 that describe the cognitive thought and action processes of en-route ATCOs in the CD&R tasks. The charts have been shaped based on extensive literature research and observing professional ATCOs at work. By focusing on MUAC, the tasks are linked to their currently operational (interface) tools. Expanding upon the work in Chapter 3, the expected impact of delegating flights and potential mitigation measures inspired by current procedures and tools are discussed. The models are then validated and quantified through a human-in-the-loop experiment (Section 4.4), for which the results are presented and discussed in Sections 4.5 and 4.6, respectively. Ultimately, the models are expected to be of use in designing human-automation flight allocation strategies for future shared airspaces.

4.2 Concept of operations

This chapter, like the rest of the thesis, takes the operations at MUAC as a baseline. MUAC is a cross-border air navigation service provider (ANSP), directing flights between 24,500 ft and 66,000 ft over Belgium, Luxembourg, the Netherlands and the western part of Germany. ATCOs work together in pairs, consisting of an executive controller (EC) and a coordinating controller (CC). The EC is responsible for all tactical control and is in direct contact with pilots, while the CC communicates with adjacent sectors and prepares the traffic for the EC. This study focuses on the work of the EC.

Unlike in flight-centric operations, ATCOs are (initially) assumed to maintain responsibility over a geographic area, in which some flights are delegated, to ease implementation in the current ATM system. The ATCOs are ultimately responsible for all flights in this area, including those delegated to the automation, and are therefore capable of regaining control at any moment over any flight.

The ground-based automation envisioned here can autonomously ensure separation between flights and issue clearances towards their planned exit point and flight level, corresponding to Level 5 from SESAR's LOA taxonomy (SESAR Joint Undertaking, 2020, p. 24). The use of simple rule-based algorithms that mimic the way ATCOs work increases acceptance and reduces the need for (complex) automation decision transparency (Westin et al., 2016a).

Despite future implementations of 4D time-based operations, potentially leading to less conflicting traffic, flights may still need to deviate from negotiated 4D trajectories due to unforeseen events such as weather or emergencies (Corver and Grote, 2016). In a similar manner, automation is not expected to actively direct flights into conflict with human-directed flights, but mixed conflicts cannot be excluded. There are various possibilities regarding solving such conflicts, sorted here by increasing LOA:

1. The ATCO has to manually resolve the conflict, by directing the flight under their control around the computer-directed flight, or delegate the flight to make it a fully automation-directed conflict (Chapter 3).

2. Automation proposes a solution to the ATCO. This can be either implemented as managed-by-exception (i.e., the proposal is automatically executed unless the ATCO rejects it within a specified time), or managed-by-consent (i.e., the proposal is only executed after the ATCO explicitly accepts it). Research indicates, however, that ATCOs are reluctant to accept decision-making aids (Bekier et al., 2012). The proposals can be limited to the delegated flights only or also involve manual-directed flights if that solution appears to be more efficient. The latter should be implemented as managed-by-consent, to give the ATCO full control over manual-directed flights.

3. Automation solves the conflict, by directing the flight under its control around the human-directed flight (Prevot et al., 2005, Strybel et al., 2016). While ATCO workload can be lowered by automatically solving conflicts, limiting the resolution to only one of the involved flights can lead to suboptimal resolutions.

It is important to stress that manual-directed flights are not necessarily excluded from all forms of automation. The current-day practice of automating most of the information acquisition and analysis stages, as well as adopting various conflict alerts is also followed here. Manual- and automation-directed flights differ mainly in the decision selection and implementation authority.

Finally, controller-pilot data link communications (CPDLC) are increasingly supplementing or replacing traditional voice-based radio transmissions (R/T). While the combined use of CPDLC and R/T has some advantages, such as sending clearances over either channel in parallel, a more distant future is being considered in which all agents communicate through CPDLC only. The complete termination of R/T provides both automation and the human with equal communication capabilities, until text-to-speech and speech-to-text technology has sufficiently evolved to close the gap.

4.3 Task analysis of common processes

This section introduces flowcharts for the CD&R processes, in which the various steps are identified and linked through letter-coding to the accompanying interface elements and tools that MUAC currently provides to its ATCOs (Table 4.1). At other ANSPs, many of these tools or variations thereof will also be available. The analysis by the authors, including a subject matter expert, is primarily based on observing two ATCOs on duty at MUAC for two hours each and simultaneously discussing their thought and action processes with them (when their workload allowed). While the exact sequence, inclusion of steps, and usage of tools can differ per person and situation, the analysis aimed to capture the most common flow of steps as executed at MUAC. It is further supported by in-text references to existing literature.

Table 4.1: MUAC support tools and interface elements.

	Tool or element	Information or action
Ⓐ	Plan view display	Horizontal flight positions and directions
Ⓑ	STCA	Characteristics of short-term conflicts
Ⓒ	Flight label	Actual and cleared level, vertical trend, next point/heading and speed (optional)
Ⓓ	FIM	Digital flight strip: aircraft type, destination, cruise level etc.
Ⓔ	VERA	Horizontal conflict verification and geometry
Ⓕ	Velocity leader length	Position extrapolation
Ⓖ	Clearance menu	Input/uplink clearances
Ⓗ	NTCA	Near-term conflict alert and probe

Color codes are used to indicate at what level of the skill, rule, and knowledge (SRK) taxonomy (Rasmussen, 1983) each block in the flowchart is executed (Table 4.2). Skill-based behavior is mostly associated with repetitive tasks or information processing that is readily available and pertains to tasks that can be instantly executed. Rule-based behavior reflects decision and action processes based on fixed rules or experience. When a new situation is encountered, knowledge-based behavior comes in sight requiring the most cognitive effort (and time). In practice, en-route ATCOs report that only few situations require this highest level, even in non-routine traffic situations (Dittmann et al., 2000). Multiple colors in a block indicate a situation-dependent level.

Table 4.2: Skills-rules-knowledge taxonomy-based color coding, plus an example.

SRK	Example
Skill-based	Comparing flight levels
Rule-based	Applying routine solutions
Knowledge-based	Generating new solutions

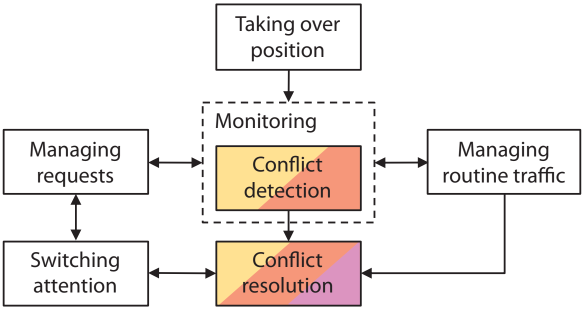


Figure 4.1: Connections between processes, adapted from Dittmann et al. (2000). See Table 4.2 for color coding.

Finally, the potential impact of delegating part of the traffic to an automated agent is discussed qualitatively for each of the processes based on the CONOPS from Section 4.2. It was preliminary tested in the simulation experiment from Chapter 3 where six ATCOs could dynamically delegate individual flights to and from an automated system. When available, examples from similar situations in current-day operations are given as a first hint towards potential solutions to reduce the impact.

4.3.1 **Monitoring**

At the start of a shift, an ATCO takes over from a colleague and receives a short briefing about any specialties (such as weather or active military areas) and flights that might require extra attention or that have been re-directed to solve a conflict. The takeover lasts not longer than one or two minutes, in which the ATCO creates an initial mental picture and sector plan. After assuming responsibility, the new ATCO enters a monitoring process that continues for the remainder of the shift. Monitoring involves updating the mental picture and sector plan, and in turn triggers all of the other processes as visualized in Figure 4.1. While the use of flowcharts may suggest purely linear processes, constant attention switching means that the processes can be interrupted or resumed due to shifting priorities.

4.3.2 **Conflict detection**

ATCOs start looking for conflicts while flights are approaching their sector, still under control of the previous sector. When a pilot calls in on the radio of the receiving sector, the ATCO first has to locate the flight, which is made easier by a radio direction finder that shows a circle around the transmitting flight on the plan view display (PVD, A). Most conflicts get identified and solved at this initial contact (Hoc and Carlier, 2002). While acknowledging that conflict detection is mostly pattern-driven when assuming flights, the focus in this chapter is on the detection of intra-sector conflicts. Throughout their shifts, ATCOs frequently check for these conflicts that may have developed well after assuming a flight. Even when multiple flights are involved, ATCOs tend to solve conflicts pair-wise (Kirwan and Flynn, 2002) and thus perform conflict detection on flight pairs, as visualized in Figure 4.2. For any given flight pair, the flowchart can be traversed via one of the 'paths' mapped in Table 4.3.

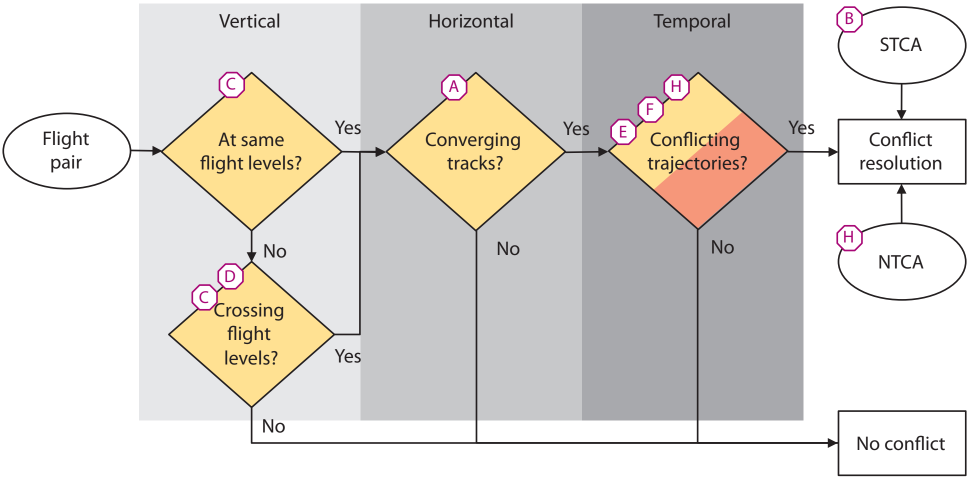


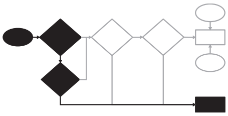
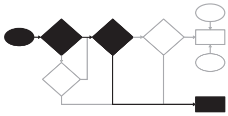
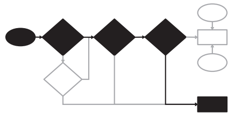
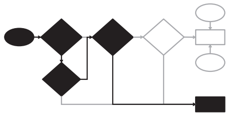
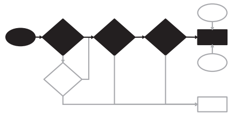
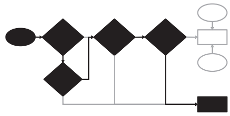
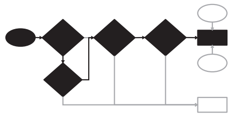
Figure 4.2: Flowchart of the conflict detection process. For letter and color coding, see Tables 4.1 and 4.2.

ATCOs first look at conflicting flight levels, as shown in the flight labels Ⓒ, before considering the directions of these flights on the PVD (Fothergill and Neal, 2013, Rantanen and Nunes, 2005). Professional ATCOs have been shown to filter out flight pairs that are safely separated in altitude from their scan patterns (Kang and Landry, 2015). The use of odd/even flight levels for traffic in 180° heading bands simplifies this task considerably, as level flights can then only be in conflict with flights from a subset of directions or with flights changing altitude (Hoekstra et al., 2016). In sectors with clear traffic streams, such flight level allocation enables further filtering of flights to be considered.

Flights changing altitude often require more effort and attention than level flights for two main reasons. First, their trajectories are harder to extrapolate due to the potential ground speed variations with vertical changes as well as possible wind conditions at different altitudes. Second, their flight levels cross with more flights that consequently have to be considered. This is especially true for flights that need to climb to their cruise level as indicated in the Flight Information Management window (FIM, Ⓓ) and subsequently descend to the coordinated transfer flight level (TFL) before leaving the sector.

Fortunately, ATCOs can often rely on their so-called 'conflict possibility library', containing hot spots within the sector where conflicts frequently occur. With the introduction of free route airspace (FRA), where flights are no longer required to follow fixed routes, the usability of such a library has diminished (Renner et al., 2018), although the majority of traffic will still follow predictable routes. For efficiency reasons, airlines prefer direct routes, increasingly made possible by FRA. Direct routes also have an advantage for the ATCOs, as it simplifies the (horizontal) detection task to a mere checking of the crossing angle between two tracks: diverging tracks will never lead to a conflict (unless involving a turning outbound flight), while parallel tracks can only be a conflict when they are already spaced less than 5 NM apart. Detection complexity is largely dependent on the convergence angle between the tracks of two conflicting flights, with shallow angles being harder to detect than perpendicular tracks (Hilburn, 2004).

Table 4.3: Possible paths to traverse the conflict detection flowchart from Figure 4.2.

Vertical	Horizontal	Temporal
 A	 B	 D
	 C	 E
		 F
		 G

Only when both vertical and horizontal separation are questionable, relative speeds are taken into account to assess whether the trajectories will actually conflict in time (Loft et al., 2007). Processing speed information requires more effort than altitude and heading (Rantanen and Nunes, 2005), potentially eliciting rule-based behavior. If the ATCO suspects an imminent conflict, the verification and advice tool (VERA, Ⓔ) can be used to validate and judge the criticality of the conflict. After selecting two flights, it extrapolates their positions along their current tracks to predict the time till and the minimum distance at the closest point of approach (CPA) between two flights. VERA only considers horizontal separation though; the ATCO needs to take the vertical aspect into account as well as any potential speed or heading changes. Any flight pair on which VERA is applied is added to a special on-screen list showing its parameters, until the ATCO cancels it. From this list, the combined PVD of both flights can be evoked, with their extrapolated positions and the corresponding CPA time. The list can serve as a to-do list or to ease the monitoring of the evolution of a conflict over a prolonged time. Apart from VERA, the length of the velocity leaders Ⓕ can also be extended (from 1 minute to 2, 4 or 8 minutes) to quickly extrapolate the future positions of flights and gauge the CPA.

A more advanced tactical prediction system is the Near-Term Conflict Alert (NTCA, Ⓗ), which extrapolates the future position along the flight's cleared route[1]. If a flight deviates from its route, or is flying on a heading, NTCA resorts to simply extrapolating the track for that flight, alike VERA. In contrast to VERA, NTCA is triggered automatically. If a loss of separation (LOS) is predicted within 4, 6 or 8 minutes (depending on ATCO selection),

[1]NTCA was not available in any of the experiments in this thesis.

an orange diamond is shown in the labels of the conflicting flights to alert the ATCO. By placing the mouse over this diamond, a VERA-like conflict geometry is shown on the PVD. Additionally, the NTCA logic is also utilized in what-if tools, which allow for the probing of alternative flight levels and/or headings before executing the pertinent clearance(s).

As a last safety measure, if a conflict is overlooked, a solely radar-based Short-Term Conflict Alert (STCA, Ⓑ) triggers 2 minutes before a LOS in the form of a red/yellow flashing radar position symbol and a yellow border around the callsign in the label. The STCA is accompanied with an entry in the Conflict Alert Message (CAM) window on the screen, showing whether the two involved flights are climbing, level or descending, and the current and predicted minimum distances between them. Both NTCA and STCA can trigger the ATCO to switch attention to conflict resolution, with STCA naturally requiring an immediate response.

Impact of flight delegation

Firstly, ATCOs have been shown to pay less attention to flights not under their (manual) control and not updating them as frequently in their mental model, potentially leading to slower CD&R (Metzger and Parasuraman, 2001). While conflicts between automation-directed flights will be automatically solved, and thus require little monitoring when the ATCO has sufficient trust in the system, the associated attention reduction can have a detrimental effect on the resolution of mixed conflicts (Chapter 3). In such conflicts, the ATCO can be taken by 'surprise' and may need to revisit the delegated flight(s) to update their mental image. In current-day operations, the CC can flag potential conflicts for the EC by adding them to the VERA list. It could be beneficial if the system acted similarly for automation-directed flights that conflict with manual-directed flights to timely inform the ATCO and reduce the risk of surprises.

Secondly, as ATCOs are not actively involved in delegated flights, they would not be updated either about (route) changes that the automation issues to flights in their sector. Something similar happens when ATCOs 'skip' flights passing through an empty part of their sector, meaning that the next sector already takes control over the flight. This is, however, only done when the flight is clear of any other traffic. With less strict conditions for delegating flights to the automation, it is even more important for ATCOs to maintain an up-to-date mental model of these flights that may interact with theirs. In current-day operations, the CC can, under certain restrictions, use CPDLC to uplink a clearance to a flight to relieve the EC. The uplink action is shown on the EC's screen by highlighting the flight's corresponding label item in magenta for a short time. Since the EC and CC are sitting next to each other, they can easily coordinate such actions.

If the other agent is an automated system, (complex) visualizations may be introduced to pro-actively communicate its actions and intentions (Jans et al., 2019), at the trade-off of increased mental demand (Wright et al., 2016). The use of a smart allocation strategy is hypothesized to reduce the need for these features by keeping ATCOs naturally in-the-loop, such that the aforementioned label item highlighting might be sufficient in most situations.

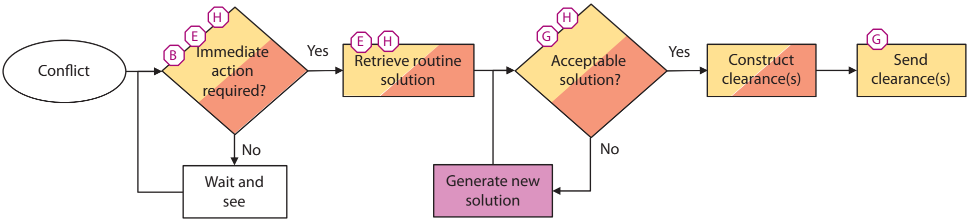


Figure 4.3: Flowchart of the conflict resolution process. For letter and color coding, see Tables 4.1 and 4.2.

4.3.3 Conflict resolution

Once a conflict has been detected, the ATCO enters the conflict resolution process, shown in Figure 4.3. Based on the criticality of the conflict – which can be derived from the time/distance to go and minimum separation given by VERA Ⓔ or NTCA Ⓗ – the ATCO needs to determine whether any action is required. With a potential conflict still far away, ATCOs may opt to 'wait and see' (Kallus et al., 1999). Especially in large sectors, uncertainties such as wind or clearances to other flights can make conflicts disappear over time with no additional effort required from the ATCO and pilots. ATCOs may therefore temporarily switch to a higher priority task and return to the conflict at a later stage if it still exists. However, under high workload, immediate actions are generally preferred as they relieve the ATCO from monitoring the situation over an extended period of time (Kirwan and Flynn, 2002, Loft et al., 2007). In most cases where a conflict triggers an STCA Ⓑ, ATCOs have already recognized and commenced resolving the conflict before the alert goes off, or they find the alert to be premature (e.g., when a fast-climbing flight can get into conflict with another flight (far) above its cleared flight level (CFL)). The 'wait and see' option then involves checking that the flight truly levels off at its CFL.

The conflict geometry visualizations from VERA Ⓔ or NTCA Ⓗ can help ATCOs in their resolution process. Most conflicts are routine conflicts that they have experienced many times during training and their career. ATCOs maintain an extensive mental 'conflict resolution library' with standard options to solve such conflicts that therefore require little mental (rule-based) processing (Loft et al., 2007). More challenging conflicts are those that are less common and thus require a custom solution, generated in a more demanding knowledge-based process. The process is repeated until an acceptable solution is found, which is then converted into a (series of) clearance(s) and input through the clearance menu Ⓖ. The ATCO then returns to monitoring and may revisit the flight at a later stage to make sure the issued clearances are properly followed up. Indications in the flight labels Ⓒ alert the ATCO if a flight has selected the wrong flight level, is not climbing or descending as instructed, or is deviating from its cleared route.

To solve conflicts, ATCOs can pick from several options. Speed clearances are rarely used in en-route control, as aircraft are mostly flying at their optimal speeds with very small flight envelopes margins. Exceptions are inbound flights that need to slow down at some point and for which the ATCO may receive speed requests from the next units. Preferred solutions are mostly those that optimize a flight's efficiency, by sending the flight either to an intermediate flight level closer to its exit level, or on a short-cut towards a point further down its planned route. At increased workloads, ATCOs give lower priority to such optimizations and are more prone to pick the first satisfactory solution they see (D'Arcy and Della Rocco, 2001) whereas in periods of low workload ATCOs may pro-actively re-inspect non-conflicting flights to see if they can further optimize them.

Any unsafe heading or altitude clearance, that would lead to a LOS within the next 2 minutes as predicted by NTCA Ⓗ, is automatically marked in the clearance menu Ⓖ, and the corresponding conflicting flight(s) is/are highlighted when the cursor is placed on the clearance value. Such features speed-up the decision selection process by quickly eliminating potentially unsafe solutions, thereby offering a solution space to select a safe clearance from. The acceptable safe solution varies by the ATCO's workload. Solutions should not increase the probability of follow-up conflicts (Niessen and Eyferth, 2001) and should minimize the need for further monitoring. This is especially important in complex or high workload situations when buffers are often increased as well (D'Arcy and Della Rocco, 2001, Kirwan and Flynn, 2002). Heading clearances usually do require a follow-up clearance to bring the flight back on its route (Corver and Grote, 2016), but can be helpful to prevent flights from turning into each other's path unnoticed. Furthermore, using parallel headings alleviates the ATCO from taking differential wind effects into account.

In line with conflict detection, the resolution of conflicts with large convergence angles can be difficult when both flights are already flying a direct route (Hilburn, 2004). A conflict with a 90° angle is considered easy and usually requires the slower aircraft to be directed behind the faster aircraft. Small convergence angles, on the other hand, often require larger track deviations up to a point where supplemental speed changes may be used to reduce the track deviations.

Impact of flight delegation

In current-day operations, flights under control by different ATCOs are less prone to get into conflict with each other due to the coordinating role of the CC and inter-sector agreements. When flights within a sector are delegated, the ATCOs need to confirm that any of their clearances have no detrimental effects on this traffic. It is expected to take more effort as, analogous to conflict detection, the flights that are not under manual control may have become dormant in or even removed from the ATCO's mental model. That would require the ATCO to actively attend to them to retrieve all (updated) information before issuing a resolution clearance.

A proposal-like setup, as described in Section 4.2, can assist in solving mixed conflicts by providing ATCOs with ready-made solutions. Instead of only highlighting the unsafe clearances in the menu (as NTCA currently does), a similar visualization could be used to indicate the suggested safe clearance. Then it would be important for the ATCO to promptly understand *why* exactly the system prefers that clearance over other (safe) solutions.

Regardless of the preceding, if flights can be dynamically delegated to the automation, a conflict may not need to be solved by the ATCO at all. Instead, they can delegate the manual-directed flight(s) involved in the conflict and have automation resolve it. Doing so is especially straightforward when one of the flights is already under automated control. Delegating both flights does mean that the automation may go beyond solving the conflict and issue additional clearances to the involved flights, unless the ATCO resumes control.

4.4 Empirical quantification

To objectively assess the impact of reduced SA of some (automation-directed) flights on CD&R of mixed conflicts, the traversal through each of the elements in the flowcharts can be linked to readily applicable measures. Since efficient and timely CD&R is key in ATC, temporal quantification seems a fitting choice (that is: how long it takes to go from one point in the chart to another). Hereto, this section describes an experiment in which a number of static traffic samples were presented to professional ATCOs. It is assumed that a completely parallel human-automation system has been established (Section 2.3.2), where the human operator does not need to pay attention to the blue automated flights for a prolonged time. In such a shared airspace, mixed conflicts can occur where the ATCO has not been actively involved with a flight that interacts with their flights. In the worst case, the ATCO has not seen an automated flight at all, until it suddenly poses a problem to one of the flights under their manual control. An example of this was observed in Chapter 3, where an ATCO did not spot a mixed conflict because the automated flight was emerging from *"a sea of automated aircraft"*.

Based on real life traffic samples, the experiment scenarios are constructed to simulate such worst-case 'pop-up' flights. They are simplified to reduce the number of variables and to measure individual contributions (Boag et al., 2006). This simplification does mean that the process of assuming flights and early checking for conflicts will be left out. The focus is on intra-sector CD&R with a short to medium-term time scale, in line with the presented flowcharts.

By analyzing the logged usage of tools from Table 4.1, a first estimation can be made of which path in the flowchart the ATCO follows. Additionally, eye tracking will show which other flights are considered and possibly re-visited, e.g., to reassess a no longer safe clearance. Similarly, the order of elements and their associated SRK-levels (i.e., higher levels requiring more effort and time, Rasmussen, 1986) can be validated. Although the duration of individual steps might not be directly traceable, *relative* differences in these measures between different (mixed) conflict types are hypothesized to give an objective measure for comparing future flight allocation strategies.

4.4.1 Participants

Ten professional en-route ATCOs from MUAC with age and experience as shown in Table 4.4 volunteered in the experiment. All participants provided written informed consent prior to their participation. The experiment setup and protocol were approved by the Human Research Ethics Committee of TU Delft under number 2574.

Table 4.4: Participant characteristics.

		Sector group		
	All	Brussels	DECO	Hannover
Number of ATCOs	10	5	2	3
Age, years (SD)	43.6 (7.1)	42.7 (6.7)	49.0 (1.0)	41.3 (8.3)
Experience, years (SD)	20.0 (6.5)	19.2 (5.3)	25.5 (1.5)	17.7 (8.2)

Figure 4.4: Experiment set-up with participant (left) and observer (right) positions.

4.4.2 Apparatus

The participants were seated behind a desk with a 27 inch monitor (1920 × 1920 pixels), and standard computer keyboard and mouse for inputs. The monitor showed SectorX, a TU Delft-built Java-based simulator with an interface mimicking the MUAC operational interface (see Appendix A). Aircraft labels could not be moved and panning or zooming of the radar display was disabled too. The interface elements (VERA, FIM and experiment dialogs) were positioned such that they were well clear of any flight symbols or labels to ease the proper distillation of gaze positions.

Eye tracking data were recorded using a Pupil Labs Core eye tracker that was to be worn as glasses (Kassner et al., 2014). Pupil Capture and Pupil Player (both version 3.5.1) were used to record and analyze the eye tracking data. None of the participants wore glasses that could hinder eye tracking measurements (contact lenses were allowed). The forward facing out-of-the-world camera recorded at 30 Hz, while the pupils were recorded at 120 Hz. Eight AprilTag markers were placed along the edges and corners of the screen to relate eye tracking data to pixels on the screen.

The experimenter was seated alongside the participant to control the simulator and note down observations. He monitored the eye tracking recording on a separate monitor. Figure 4.4 shows the complete test set-up.

4.4.3 Procedure and participant tasks

During the briefing, the participants were instructed that they could only control green flights. Blue flights were assumed to be controlled by a different ATCO and therefore not susceptible to receiving any clearances from the participants. Apart from the initial clearances shown in the flight labels, the blue flights would not receive any follow-up clearances. The ATCOs could practice on six training scenarios that were identical in nature to the measurement scenarios.

The participants were then asked to put on the eye tracker glasses and make sure they were in a comfortable position before a nine-point on-screen calibration sequence was performed. This calibration was repeated before each measurement block of 15 scenarios. At the end of each block, an on-screen validation sequence was performed to account for the buildup of slippage errors due to small movements of the eye tracker due to the participants' face movements. Only one additional intermediate calibration was performed between two scenarios, when one of the participant's eye tracker moved noticeably.

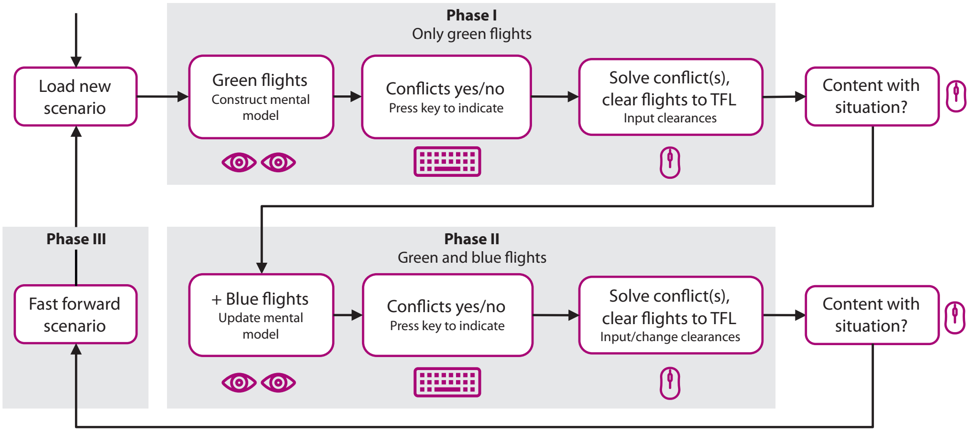


Figure 4.5: Experiment procedure, as repeated three times for all 15 scenarios.

4

Each scenario followed a standard procedure consisting of three phases, as outlined in Figure 4.5. This exact procedure was also followed in six training runs to prepare the participants for the measurement runs.

1. *Phase I: only green flights*

   A static traffic situation was shown to the participants consisting of a number of green-colored aircraft. The participants had to indicate whether they believed any flights within the sector boundaries were currently in conflict (i.e., they would get within 5 NM horizontally and 1,000 ft vertically if no clearance was issued). VERA was available at this point, but no clearances could be given just yet. As soon as the participant was confident in their assessment, they would submit it by pressing the space bar and clicking an on-screen 'Yes' or 'No' button, depending on whether they believed there was a conflict. The ATCOs were briefed to press the space bar as soon as they had spotted a conflict and to not wait until they had analyzed the whole scenario.

   Participants were then given the opportunity to issue altitude clearances to any of the flights by making use of the clearance menu from the interactive labels. Conflicts had to be solved instantaneous and could not be left for 'wait and see'. No heading, speed or trajectory clearances could be given, nor any vertical rates. Flights should preferably be cleared to their TFL, unless it was unsafe to do so. Then an adjacent level should be selected.

Once the ATCO was sufficiently content, they could press a 'Done' button to move on to Phase II.

2. *Phase II: green and blue flights*

   One or two blue-colored flights were revealed alongside the existing green-colored flights. The participants had to re-judge this new situation and once again indicate whether any or none of the green flights was in conflict (either with another green or a blue flight) as soon as they could. Some of the conflicts would have been solved already in Phase I, while new conflicts may have been introduced by the blue flights.

   Participants then issued altitude clearances as before, with the notion that clearances could only be given to the green-colored flights. It was possible to overwrite clearances from Phase I, if required by the new situation. Once the ATCO was content with the given clearances, they pressed the 'Done' button again.

3. *Phase III: fast forward*

   The scenario was unfrozen and fast forwarded at 50x real time for 5 minutes of scenario time (6 seconds real time), to give the ATCOs a chance to see the implication of their actions. This allowed them to observe whether their clearances had the desired effect and provided a break between scenarios. After 6 seconds, the simulation was frozen again and the ATCOs could move on to Phase I of the next scenario by pressing a 'Next scenario' button. Before doing so, the ATCOs were told to make sure they were ready to immediately start the conflict detection process, as the measurements would start promptly.

This procedure was repeated for a consecutive block of 15 scenarios, after which the ATCO could briefly relax before re-calibrating the eye tracker for the next block. After three of such blocks (45 measurements in total), the experiment was concluded.

4.4.4 Scenario design

The scenarios consisted of two to five flights. The number of flights was relatively low, to streamline the collection of a sufficient number of situations in limited time. All scenarios featured the same artificial octagon-shaped sector of 80 by 80 NM, with four waypoints (one in each cardinal direction) acting as exit coordination points (XCOP). An artificial sector was chosen to avoid any sector group specific advantages and to be able to transform the scenarios to prevent recognition (see Appendix C). The following situations could demand ATCO input on green flights:

- Two flights with a predicted loss of separation (conflict).
- The flight still needs to climb/descend to its transfer level, in which case:
  - There may be a clear direct path to this TFL, or
  - One of the other flights may block the TFL, requiring an intermediate level-off.

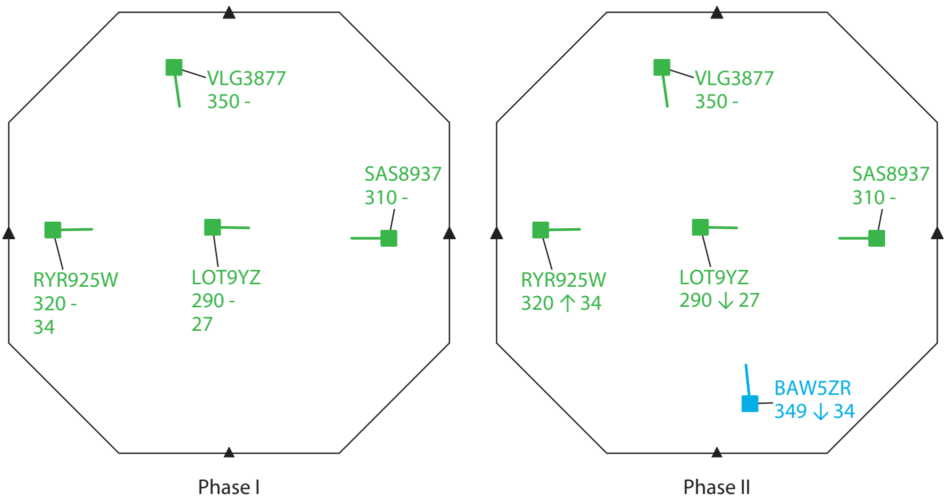


Figure 4.6: Scenario 12 in two phases (symbols and labels not to scale).

As an example, Scenario 12 is shown in Figure 4.6. Phase I is conflict free, with all six flight pairs eliciting conflict detection flowchart Path A (Table 4.3). Two flights, RYR925W and LOT9YZ, not yet meet their exit conditions (third line of the label) and require, respectively, a climb and descend clearance. After issuing these clearances and proceeding to Phase II, the newly introduced BAW5ZR creates four additional flight pairs. Three of which are conflict free (all Path A), and one is a conflict (Path G). The conflict invalidates the clearance issued to RYR925W, necessitating an intermediate level-off at FL330.

Table 4.5 gives an overview of which paths from the conflict detection flowchart in Figure 4.2 were present in each scenario and phase. Note that each path is hypothesized to have an associated cognitive load, meaning that different combinations of the same number of paths can require varying ATCO effort. Conflict-free and action-free scenarios were also included to raise the ATCOs' alertness. The following design rules were applied to scope the experiment and prevent confusion about procedural discrepancies:

- All flights were on direct routes to their XCOP: one of the waypoints.
- If two blue flights were in conflict with each other, they would already be cleared to another flight level to solve the conflict, as ATCOs were not able to control those flights. These clearances were visible in the flight labels.
- Conflicts could not involve more than two flights, as multi-flight conflicts are relatively sparse in real-life and ATCOs tend to approach these pairwise.
- All flights were of the same Airbus A320 aircraft type.
- Data associated with each flight were kept to a minimum (e.g., no destination airports or requested cruise levels).

All scenarios were presented three times (referred to as a, b and c), either rotated, mirrored and/or moved up/down in flight levels, to get three measurements per scenario without participants recognizing this repetition (see Appendix C).

Table 4.5: Number of flight pairs per scenario and phase with corresponding conflict detection paths (Table 4.3).

	Scenario, phase																															
	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15		Total	
Path	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II	I	II
A	2	4	2	8	3	5	2	7	2	5		7	1	3	2	4	2	4			2	5	6	9		1	1	5		1	25	68
B			1	1							1	1																			2	2
C						1			1	1																					1	2
D								1	1	1	1	1					1	2		1	2	2								1	5	9
E	1	1									1	1	1	1							1	1			1	1					5	5
F				1	2	3	1	1	2	2			1	1	4	5			1	1	1	1							1	1	13	16
G		1			1	1		1		1				1		1			2	4		1		1		1		1			3	14
	3	6	3	10	6	10	3	10	6	10	3	10	3	6	6	10	3	6	3	6	6	10	6	10	1	3	1	6	1	3	54	116

4.4.5 **Independent variables**

There was one independent variable in the experiment: the scenario, as described in Section 4.4.4. The scenarios were directly related to the conflict detection flowchart from Figure 4.2 and varied in number of flights and conflict characteristics (if present).

4.4.6 **Dependent measures**

The following measures were either collected in real-time or derived post-hoc from eye tracking data:

- Conflict detection status as submitted by the ATCO and its timing,
- Timing of VERA inspections and associated flight pair(s),
- Timing and flight level of altitude clearances,
- Eye tracking data:
    - Dwell time, the total time spent looking at each flight and flight pair, and
    - Gaze transitions, the number of times ATCOs alternated their gaze between flights, after looking at another flight for at least 100 ms (as used by, e.g., Nordman et al., 2023).

4.4.7 **Data analysis**

Eye tracking data, in combination with on-screen markers, allowed for the distillation of gazes on the radar screen. As the experiment involved static and visually well-separated (non-overlapping) flights, it was deemed unnecessary to look at fixations. To correlate gazes to specific flights or interface elements, the radar display was divided into areas of interest according to the Voronoi tessellation method, which determines for each location which seed (object of interest) is closer to it than any other seed (Voronoi, 1908). Using such partitioning caters for potential accuracy issues by assigning all gazes in a wider area to a certain object. An example for Scenario 12 is given in Figure 4.7, where a seed was placed at each of the following elements to create the Voronoi polygons:

- Radar position symbol (blip),
- Center of flight label,
- Tip of velocity leader,
- Four corners of the dialogs (VERA, FIM and experiment).

Because the flowcharts from Section 4.3 deal with the assessment of flight *pairs*, the dwell time measures of *individual* flights need to be transferred to flight pairs. For this, transitions between flights (gaze moving from one flight to the other, Kang and Landry, 2015) are matched to the respective flight pair and its corresponding flowchart path from Table 4.3. As it is impossible to reliably determine how much of the gaze on a flight pertains to the flight that is looked at before or after it, it is assumed that a flight's dwell time is proportionally devoted to the directly preceding and following flights. Figure 4.8 illustrates this concept. Note that the dwell times of the first and last gazed flight in a scenario, that lack a preceding or succeeding flight, are completely associated to just one pair.

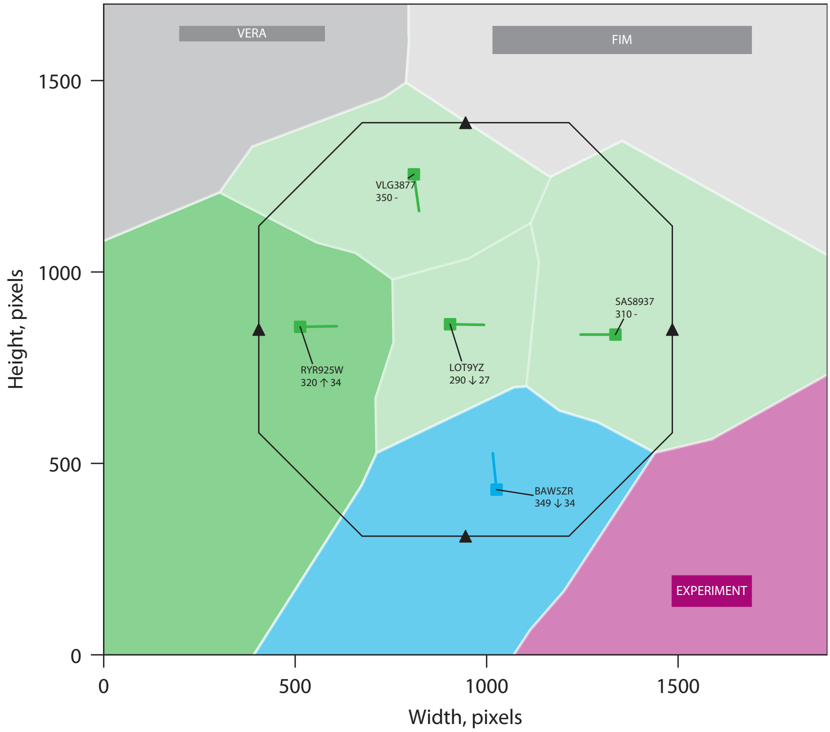


Figure 4.7: Area of interest classification through Voronoi tessellation for Phase II of Scenario 12. Flights RYR925W and BAW5ZR are in conflict, as indicated by the darker color of their area of interest.

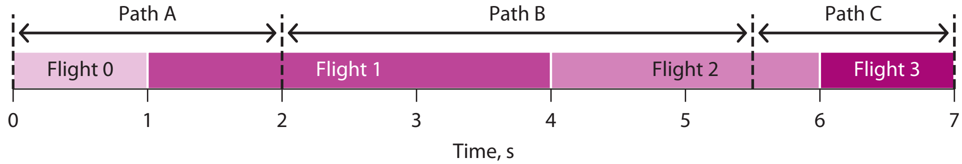


Figure 4.8: Calculation of dwell time per flight pair and corresponding conflict detection flowchart path.

4.4.8 Hypotheses

The following hypotheses were formulated:

H1 Conflict detection time is proportional to the number of flights appearing on the screen.

H2 Sequential steps in the conflict detection task, as outlined in Figure 4.2, require increasing cognitive effort. Vertical checks are the fastest, followed by directional and finally temporal overlaps.

H3 ATCOs are faster in detecting a conflict in Phase II when it involves a green flight for which they had given a clearance in Phase I (e.g., as shown in Figure 4.6), compared to cases where they did not interact with that flight.

4.5 Results

4.5.1 Conflict assessment

For every scenario, the ATCOs first had to indicate whether they believed there was a conflict. Every block in Figure 4.9 corresponds to one repetition of a particular scenario for a single ATCO and the block's color indicates the submitted conflict assessment. Each ATCO is represented by one row of blocks, while each column corresponds to one of the three repetitions of a scenario (a, b and c).

Overall the ATCOs agreed on whether there were conflicts present, as shown by the generally consistent colors per scenario. An outlier in this regard is ATCO 4, who consistently flagged all flights that would cross another flight's path if they were to be cleared to their TFL, despite being briefed to only consider the current and cleared flight levels. Furthermore, Scenario 9 stands out as ambiguous in Phase I, with six ATCOs changing their mind for one or two of the repetitions. In this scenario, two level flights were on converging paths with a CPA separation of 13 NM, which does not strictly breach the separation criterion of 5 NM. It may have been considered unsafe by the ATCOs because, in reality, flights that follow a route instead of a heading may deviate from their route or make unanticipated heading changes, which can lead to a decrease in CPA separation.

In 22 occasions (5% of 450 assessments), ATCOs refrained from issuing the intended clearances in Phase I, giving them a potential 'advantage' in Phase II over the other ATCOs, as they would not encounter a designed conflict. This is indicated in Figure 4.9 with dots and explains some of the 'incorrect' conflict assessments in Phase II, such as ATCO 2's solitary conflict-free assessment of Scenario 12b. Due to the small number of such occurrences, and given that the ATCOs sometimes corrected their own 'mistake' in Phase II, these results are *not* excluded from the remainder of the analysis.

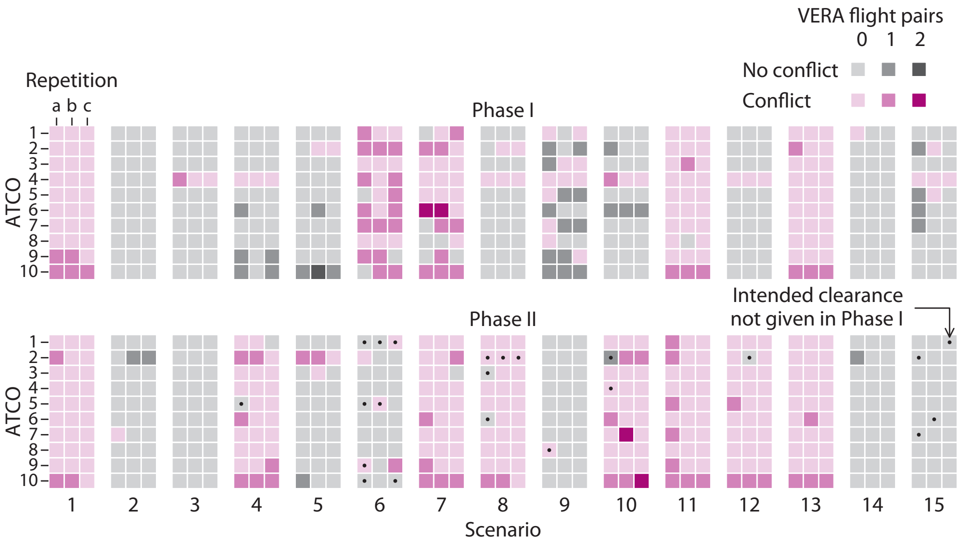


Figure 4.9: Conflict assessment and VERA usage per scenario repetition (a, b and c) and ATCO in both phases. Colors show the conflict status that each ATCO submitted, while the shade of that color indicates whether VERA was used, and if so, on how many flight pairs. Dots indicate that an intended clearance was not given in Phase I.

4.5.2 **VERA support tool usage**

Figure 4.9 also shows the ATCOs' use of VERA to measure the predicted CPA between two flights and/or show their conflict geometry. Scenarios 6, 7 and 9 stand out with more VERA usage than others in Phase I. Scenario 6 involved two flights that would have had a CPA of 8.4 NM when flying at identical speeds, but due to their different speeds the actual CPA was only 4.5 NM. For Scenario 9, the high VERA usage is reflected in the relatively inconsistent conflict assessment, with four ATCOs only flagging it as a conflict when they did not use VERA, whereas they flagged it as conflict-free when they did use VERA.

To ease further analysis, Table 4.6 lists the aggregated numbers per ATCO and scenario repetition. Overall, the VERA tool was used on 75 and 51 flight pairs in Phases I and II, respectively. A learning effect is observed in the decreasing number of VERA invocations with subsequent occurrences of the same scenario in both phases. The largest decrement for all ATCOs combined was between the first and second repetition (a and b), which flattened out in the third repetition (c). This greatly varied between the ATCOs though, with some ATCOs showing a strong decrease over the repetitions (e.g., ATCO 6) while other ATCOs were more consistent (e.g., ATCO 10) or did not use VERA at all (ATCO 8). Of the checked pairs, 40 were in conflict (CPA < 5 NM and crossing flight levels), while six others were not currently in a conflict, but would be conflicting if they were cleared to their TFL.

Table 4.6: Number of flight pairs on which VERA was used per phase and scenario repetition.

	Scenario	ATCO										
	repetition	1	2	3	4	5	6	7	8	9	10	Total
Phase I	a	1	6	1	3	1	7	2	0	4	7	32
	b	0	2	1	0	1	4	3	0	4	8	23
	c	1	2	0	1	2	2	3	0	1	8	20
		2	10	2	4	4	13	8	0	9	23	75
Phase II	a	1	6	0	0	2	3	1	0	2	9	24
	b	0	4	0	0	0	1	2	0	0	8	15
	c	0	3	0	0	0	0	0	0	2	7	12
		1	13	0	0	2	4	3	0	4	24	51

4.5.3 **Gaze timelines**

When presented with a traffic picture, ATCOs gaze over the flights and their labels to assess the situation. The collected eye tracking data reveals these gaze patterns. An example of a gaze timeline for a single repetition of Scenario 12 is shown in Figure 4.10, where the colors correspond to the areas of interest defined in Figure 4.7.

It is clear that using the press on the space bar (vertical dashed line) or the actual submission of a conflict rating (right end of the stacked bars) were not good measures of conflict detection time. Many ATCOs either forgot to press the space bar timely (e.g., ATCO 3), or they spent considerable time thinking about whether they had to answer

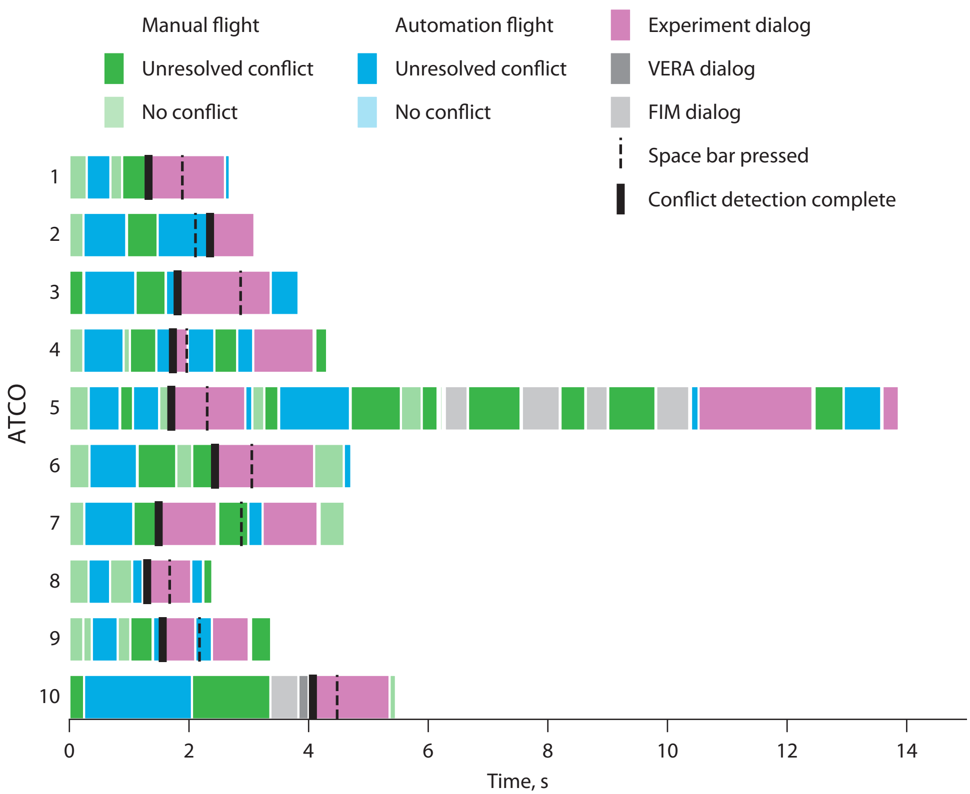


Figure 4.10: Gaze timelines of all ATCOs for a single repetition of Scenario 12 in Phase II. Colored blocks correspond to the areas of interest defined in Figure 4.7.

'yes' or 'no' to indicate a conflict. ATCO 5 tried to resolve the conflict after having clicked the conflict button without pressing the space bar first (which would have unlocked the clearances), leading to an excessively long duration of nearly 14 s. Therefore, to determine a more accurate conflict assessment duration the time is capped at the onset of the first gaze on the experiment dialog that lasted more than 200 ms, indicated by the solid black vertical line in the figure.

Figure 4.10 also shows that all ATCOs were looking at a green flight when Phase II started (Time $=$ 0 s). Within circa 300 ms, all of them moved their gaze to the newly shown blue flight, which was generally followed by gazes on one or two green flight(s) before looking at the experiment dialog. The green flight that was involved in an unresolved conflict with the blue flight clearly dominates the dwell time, as indicated by the relatively large dark green (and blue) sections. Interestingly, seven out of ten ATCOs did not look at VLG3877 at all in Phase II. As shown in Figure 4.6, this flight was flying head-on with a blue flight, albeit at an adjacent flight level and thus conflict-free. Furthermore, note that ATCO 10 was the only ATCO who used VERA in this scenario. The gaze timeline shows how this ATCO looked at the VERA dialog to confirm the CPA distance just before submitting a conflict assessment.

To account for learning effects (as observed in e.g., the use of VERA, Table 4.6) and given the time-critical nature of conflict detection, for the remainder of this analysis each ATCO's results were trimmed down to the single repetition of each set of three for which the ATCO showed minimal detection time in Phase II. In the vast majority of scenarios this was either the second (31%) or the third (61%) repetition. The largest reduction in mean detection time for all participants and scenarios combined was observed between the first and second repetition (1.0 seconds), which flattened between the second and third repetition (0.2 seconds). This flattening is consistent with the observed trend in VERA usage (Table 4.6).

4.5.4 Conflict assessment duration

Figure 4.11 shows the conflict assessment duration in Phase I. Each data point corresponds to the fastest duration of that particular scenario for one ATCO. Data points shown in black indicate that the ATCO used VERA before the measurement was capped, as explained in Section 4.5.3, while for the other points VERA may have been used later in the scenario or not at all. Scenarios are ordered according to their median values (indicated by the thick black lines) and are grouped by their number of flights, for which the gray rectangles denote the median detection time.

As expected, the number of flights had a significant effect on the measured duration, as confirmed by a Friedman's test ($\chi^2(2) = 11.4, p = .003$). Conover's pairwise post hoc comparisons with Bonferroni correction showed that the duration was significantly higher when the number of flights was doubled from two (median $= 1.5$ s) to four flights (median $= 3.6$ s, $p = .011$), with a moderate effect size ($W = 0.570$). Scenario 11 is a noticeable outlier, encompassing four flights but a duration that is comparable to scenarios with only three flights. As this was the only scenario with four flights that contained a conflict in Phase I, it seems plausible that the ATCOs did indeed stop assessing the situation further once they had spotted the conflict, like they were briefed to do. Apart from this, the variation between scenarios with an identical number of flights hints at a further dependence on scenario-specific characteristics.

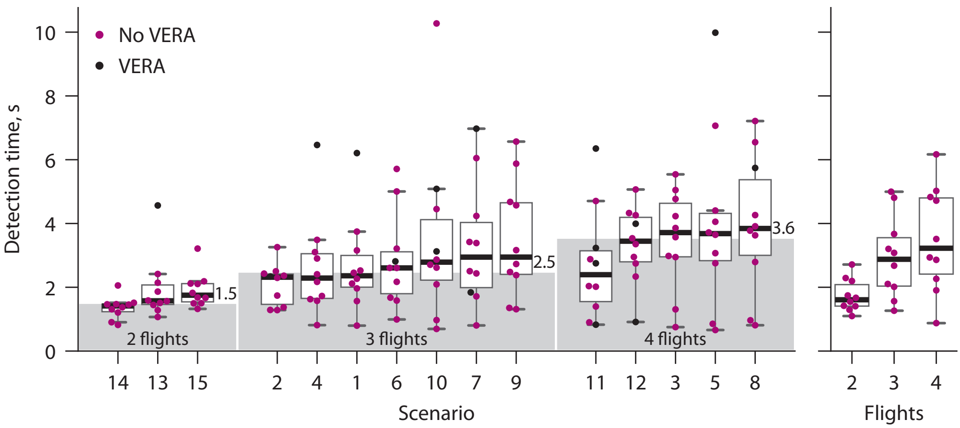


Figure 4.11: Minimum detection times in Phase I for all ATCOs, ordered by median time per scenario and grouped per total number of flights, for which the gray rectangles show the median time.

In Phase II, the total number of flights cannot be directly linked to the assessment duration, as visible in the top half of Figure 4.12. Whereas one might expect to see a similar staircase pattern as in Figure 4.11, the number of additional blue flights seems to have a stronger impact than the total number of flights (lower half of Figure 4.12). Two instead of one additional blue flight raises the median detection time from 1.7 s to 3.4 s, a significant increase ($t(9) = -8.379, p < .001$) with a large effect size according to Cohen's d (-2.65). This suggests that the detection and resolution loops in Phase I have already ruled out several flight pairs that do not need to be revisited in Phase II. Furthermore, the composition of the individual flight pairs might explain some of the variance found between scenarios with similar number of flights, just like in Phase I.

Scenarios 1, 8 and 11 were very quickly traversed by almost all ATCOs. In all three scenarios, a green flight cleared to its TFL in Phase I was unable to safely reach the CFL due to a blue flight blocking that level. The ATCOs quickly recognized these conflicts, as shown by relatively low detection times (similar to Scenarios 13 and 15 with just two flights) and small variance between ATCOs. Scenario 10 is found at the other end of the spectrum. In many cases this scenario required two previously issued clearances to be modified, due to a blue flight blocking the TFL of two green flights. While this is an easy conflict to spot, the solution is more difficult, as now three flights are involved. This is reflected in the above-average detection time and may indicate that ATCOs simultaneously started to (mentally) resolve the conflict.

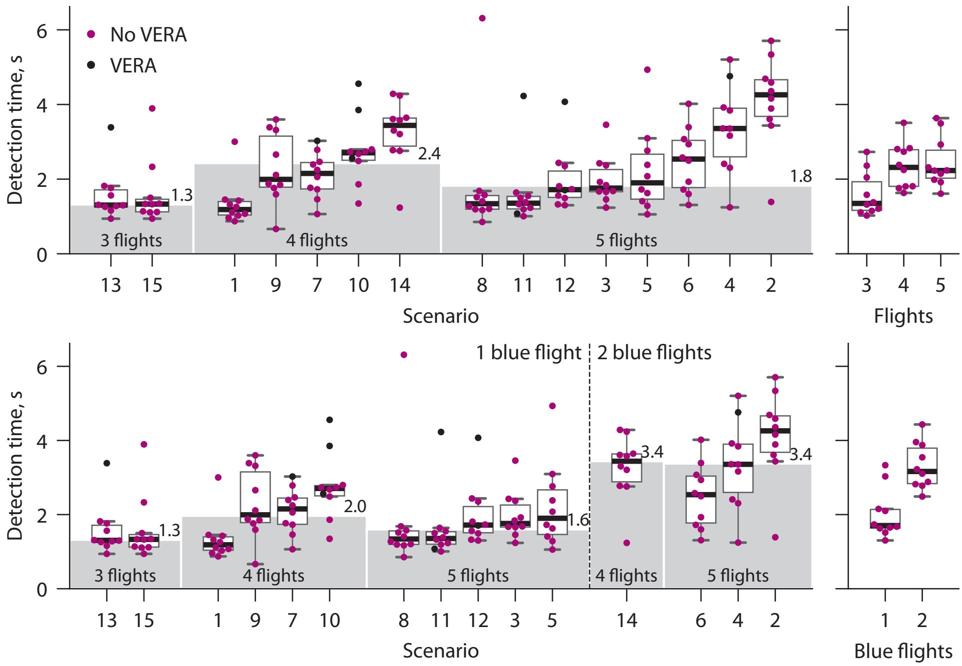


Figure 4.12: Minimum detection times in Phase II for all ATCOs, ordered by median time per scenario and grouped per total number of flights, for which the gray rectangles show the median time. The bottom half of the figure shows the same data as the top half, but with an additional grouping per number of blue flights.

4.5.5 Visual attention per flight (pair)

As seen in Figure 4.12, the number of blue flights can explain some of the differences in detection duration. However, several scenarios with an identical number of green and blue flights showed considerably different detection times, such as Scenarios 1 and 10. To get a more detailed analysis of which specific flight(s) (pairs) contributed most to the conflict detection duration, the visual attention per flight was determined in terms of total dwell time and number of gaze transitions (i.e., moving gaze from one flight to the other, with a minimum dwell time of 100 ms).

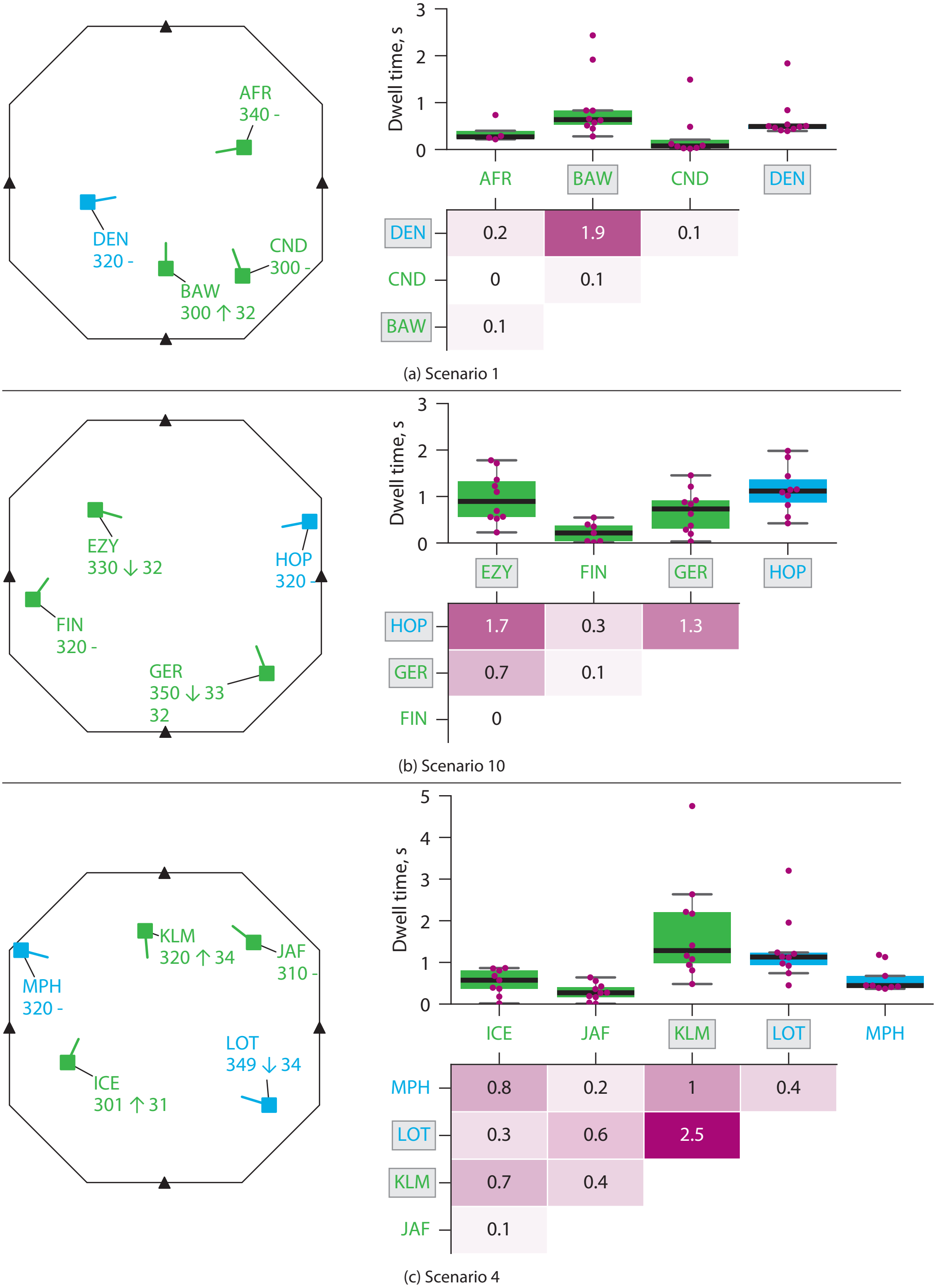
Figure 4.13 shows the results for three characteristic scenarios, selected on the basis of their detection durations. For each scenario, the left-hand side of the figure shows the traffic situation at the start of Phase II, provided that the ATCOs had issued the expected clearances in Phase I. Note that the callsigns as shown here have been simplified from those presented in the experiment to ease the discussion, and aircraft symbols and labels have been up-scaled for legibility. The right-hand side of the figure shows each ATCO's dwell time per flight (top) and a matrix of the average number of gaze transitions per ATCO for any flight pair (bottom). The callsigns of conflicting flights that require ATCO interference are outlined with rectangular boxes.

Starting with Scenario 1 (Figure 4.13a), BAW had been cleared by the ATCOs in Phase I to climb to FL320, a flight level that is now blocked by DEN. In line with the relatively low overall conflict detection duration for this scenario (median of 1.2 s, Figure 4.12), all flights showed low dwell times, but the two conflicting flights have significantly higher dwell times than the other flights. This higher level of attention for the conflict pair is confirmed by the number of gaze transitions, which is also the highest of all pairs.

Scenario 10 had a significantly higher detection duration in Figure 4.12 than Scenario 1 ($t(9) = -8.027, p < .001$), despite an identical number of flights. This scenario, shown in Figure 4.13b, was a special case as it involved two green flights that needed adjustment in Phase II. Both EZY and GER had identical exit levels of FL320. Since this would result in a conflict, the ATCOs cleared EZY to FL320 in Phase I, while GER was initially cleared to an intermediate FL330. The introduction of HOP at FL320 in Phase II created a conflict with EZY, necessitating this flight to stay at FL330, which then became unavailable to GER. Similar to Scenario 1, the ATCOs spend most of their visual attention on these three interacting flights. FIN was barely looked at, despite the fact that it was on the same flight level as HOP (but horizontally separated).

As a final example, Scenario 4 contained three green and two blue flights. As shown in Figure 4.13c, the ATCOs quickly established that MPH was not conflicting with 'their' flights, leading to one of the lowest dwell times of all flights in this scenario. The other blue flight, LOT, was descending to the same flight level as KLM and thus required the ATCOs to adjust their clearance for this flight. The two flights of this mixed conflict had the highest dwell times in the scenario and the gaze transition matrix shows that the ATCOs' gaze mainly jumped between these two flights.

The other scenarios showed similar results. In general, the flights with the longest dwell time were often also part of the flight pair that received the highest number of consecutive visits. This always included a blue pop-up flight when it created a conflict.



(a) Scenario 1

(b) Scenario 10

(c) Scenario 4

Figure 4.13: Per-flight dwell times and mean number of gaze transitions between flights for three scenarios in Phase II. Flights with ‘boxed’ callsigns are in conflict.

4.5.6 Visual attention per flowchart path

Taking the visual attention analysis one step further, each flight pair can be linked to a corresponding flowchart path (Table 4.3) based on the required steps to assess the pair's conflict status (i.e., first checking for sufficient vertical, then horizontal and finally temporal separation). Following the method outlined in Section 4.4.6, the previously discussed dwell times per flight and gaze transitions between flights were consolidated into dwell times per flight *pair*.

Figure 4.14 shows the resulting dwell times, averaged per ATCO and split per flight pair type. Note that each type was represented between 0 and 40 times per path (Table 4.7), such that only Paths A and G were represented by all types. Paths C, D, and F did not occur with fully automated pairs, while Paths B and E only occurred in the form of manual pairs. This incomplete representation of types makes it impossible to do a proper statistical between-paths comparison and to test the hypothesized increasing cognitive effort with increasing traversal depth of the flowchart.

Overall, the median dwell time was higher for flight pairs that involved a blue 'popup' flight (i.e., mixed or automated flight pairs) than for manual flight pairs, irrespective of the path. In addition, fully automated flight pairs evoked dwell times that were, compared to mixed pairs, higher for Path A and lower for Path G. For seven ATCOs the Path C duration

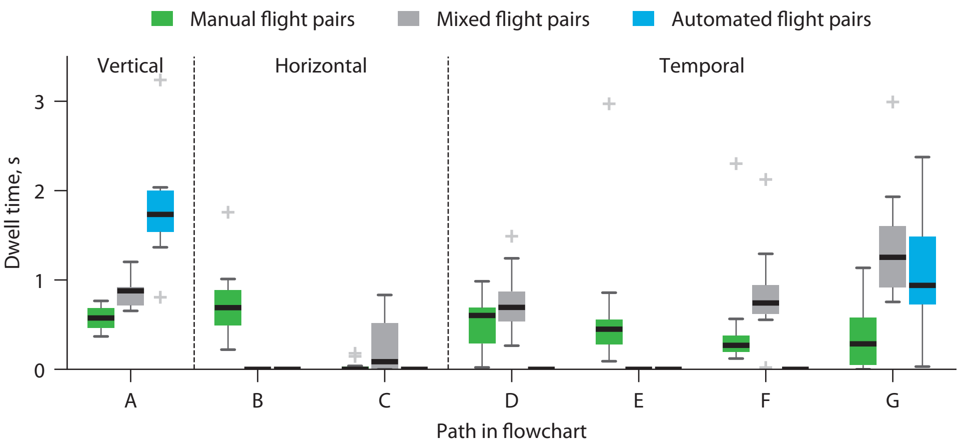


Figure 4.14: Average dwell time per ATCO for each conflict detection flowchart path (Table 4.3) in Phase II, split per flight pair type. The headers above the figure refer to the depth of traversal through the flowchart (Figure 4.2), hypothesized to evoke increasing cognitive effort from left to right.

Table 4.7: Number of occurrences of each flowchart path per ATCO, split per flight pair type and scenario phase.

		Path in flowchart						
	Flight pair type	A	B	C	D	E	F	G
Phase I	Manual	25	2	1	5	5	13	3
Phase II	Mixed	40	-	1	4	-	3	10
	Automated	3	-	-	-	-	-	1

was zero in Phase I. The differences between the median dwell time of manual and mixed pairs is just 90 ms for Path D, making this a negligible difference as well.

Because the main research objective here was to determine the duration of transitioning a path in the conflict detection flowchart, preferably the time differences between paths are compared. Although a statistical significance could be calculated for each of these observations, Wasserstein et al. (2019) and others, reason that statistical tests are of limited value in many cases, e.g., where insufficient data is obtained. Given that some path types were only encountered in one scenario (e.g., automated pair, Path G), the specific layout of that scenario has an excessive influence on the measured dwell time, meaning that a test would not actually compare paths, but scenarios.

4.6 Discussion and recommendations

4.6.1 Hypotheses

The number of new flights that were introduced to a scenario showed a statistically significant effect on conflict detection duration, confirming Hypothesis H1. This was the case in both Phase I and II, where the latter phase showed that the total number of flights in the scene had little impact. ATCOs are able to 'memorize' traffic pictures to a great extent, especially when it comes to an aircraft's position and altitude (Gronlund et al., 1998), so only existing flights relevant to the 'pop-up flights' were revisited and received significantly more visual attention. This result supports the notion that mixed conflicts, involving flights that the ATCO may have not or only superficially inspected, can be more cognitively demanding than conflicts between manual flights.

Unlike hypothesized in H2, the presumed depth of traversal through the conflict detection flowchart of Figure 4.2 did not show a clear effect on dwell time (and therefore conflict detection duration). This can be due to several reasons. To start with, the greatly simplified scenarios required little cognitive processing. They elicited hardly any (if any) knowledge-based behavior, which was expected beforehand to lead to the biggest measurable change. Furthermore, expert operators may not adhere to the steps in a linear fashion as suggested by the presented flowcharts, but may employ shortcuts like described by Rasmussen (1986). This could potentially impact the reliability of temporal measurements, as operators may use shortcuts in some scenarios only.

Besides, not all flight pair types were represented in all flowchart paths and if they were, the number of data points was rather low. The paths resembled less than 20 flight pairs per ATCO (and as low as two pairs for Paths B and C), except for Path A which was found in 68 pairs (Table 4.5). This means that for the majority of paths the specific scenario had a large impact on the measures, such that no trends can be observed with respect to the flowchart traversal depth.

Finally, it turned out to be infeasible to determine whether the ATCOs solely performed conflict detection, without simultaneously resolving found conflicts. These two tasks cannot be entirely isolated as they are cognitively closely interlinked, as indicated by the loop around 'acceptable solution' in the conflict resolution flowchart of Figure 4.3.

The impact of interactions in Phase I on the conflict detection time in Phase II could not be reliably determined due to the limited number of scenarios in which this occurred. However, the scenarios where it did occur led to the fastest detection times in Figure 4.12, hinting at a (small) reduction effect on the duration. Based on the preceding, H3 could not be confirmed. Regardless, in an operational setting 'popup' flights can occur up to 20 minutes after a clearance has been issued (depending on the size of the sector), which may lead to more pronounced effects of manual interaction on conflict detection speed.

4.6.2 Experiment

As briefly touched upon in the previous section, the experiment design had a significant impact on the results and subsequent analysis. The ATCOs were not familiar with the presented zoom-level nor sector shape, which was reflected in some ATCOs noting that it was harder to judge distances and speeds than in their daily work. This may have led to longer detection times and a potentially higher-than-normal use of VERA. Indeed, the decrement in VERA usage over the three scenario repetitions suggests that the ATCOs became more familiar with the scale of the display as the experiment progressed.

In addition, several ATCOs considered the required press on the space bar, before being able to set a conflict status, troublesome and non-intuitive. This led to a delayed response time in some cases, especially in the first repetition block. In latter blocks this had only a minor influence and, together with the order-balancing between participants, therefore little impact on the analysis presented here. Nevertheless, using eye tracking data to determine when ATCOs had spotted the conflict proved to be more reliable and generally more convenient for the ATCOs.

Echoing the preceding discussion, future work should also consider more realistic and diverse scenarios. Special care must be taken when designing these scenarios to ensure that all flowchart paths are sufficiently present with all three flight pair types (i.e., manual, mixed and automated). This will allow a proper trend analysis which could not be performed for this chapter's experiment. Rather than isolated situations, complete traffic scenarios are expected to evoke more knowledge-based behavior and will ensure that the results are more indicative of operational CD&R time scales. However, such scenarios will be more difficult to analyze as tight experiment control is traded for realism and ATCO freedom.

4.6.3 Conflict resolution

Given the limited success of the chosen method in quantifying the conflict detection task, it does not directly provide a clear recipe for quantifying the conflict resolution task. Apart from the aforementioned general issues with quantifying (complex) cognitive processes, the present experiment setup and in particular the scenario design, resulted in insufficient resolution flexibility and variation. All conflicts in Phase I could be solved by sending flights to or towards their exit flight levels. Although this was by design to ensure consistent start-situation in Phase II, it also reduced the resolution task to a mere check whether this logical solution was safe or not. Again, more realistic and diverse scenarios are needed to provide sufficient ground for a conflict resolution analysis, at the risk of complicating the analysis.

Concluding, it might be beneficial to take a more holistic approach and combine the CD&R task into a single 'complexity' measure that could represent the entire cognitive effort required to guide a flight safely and efficiently through a sector. This will simultaneously circumvent the problematic (and probably futile) isolation of conflict detection as an independent subtask.

4.7 Conclusions

This chapter introduced flowcharts representing the common ATC tasks of conflict detection and resolution. The charts show which steps an ATCO generally traverses in order to complete these tasks. Depending on a flight pair's composition, the steps require increasing cognitive effort. An attempt to empirically quantify these steps through a simulation experiment with professional ATCOs was only partly successful, mostly due to the intrinsic complexity of cognitive processes that are not (easily) captured in linear steps. Frequent task switching and the interdependence of conflict detection *and* resolution further contributed to this. However, results indicate that the number of flights is not the main factor impacting CD&R workflows. Presumably, interactions between flights (or a lack thereof) play a more important role, next to ATCO perceptions and experiences.

Future research should identify exactly which interactions between flights, in more realistic scenarios, add to the perceived complexity of guiding a particular flight through a sector (i.e., combining both CD&R). In addition, it should be established to what level ATCOs are consistent in their perceptions of complexity. This can then be used to drive a complexity-based flight allocation scheme in future shared airspaces, that can minimize the occurrence of mixed conflicts where the authority and responsibility over each flight is assigned to a different agent (i.e., human or automation).

5

Flight-centric complexity

The classification, and subsequent allocation of flights to either human or automation, is preferably based on objective measures relating to the traffic situation. Existing complexity models are, however, often used for capacity predictions or airspace restructuring and primarily to assess the complexity of a sector as a whole. This chapter uses empirically collected flight complexity ratings from 15 professional en-route air traffic control officers. By analyzing the interactions between flights that they themselves included in their complexity assessments, a classification model is established to differentiate between basic and non-basic flights, and to identify which traffic features play the largest role. This can then serve as a starting point for an automatic allocation algorithm that distributes flights between a human controller and the automation.

This chapter has been published as:

de Rooij, G., Stienstra, A., Borst, C., Tisza, A.B., van Paassen, M.M., and Mulder, M. Contributing Factors to Flight-Centric Complexity in En-Route Air Traffic Control. In *Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)*. 2023b

Part of the data is accessible through:

doi:10.4121/f343aa29-6024-4b07-bbcd-5678f61badec

5.1 Introduction

In striving for safe and efficient operation of air traffic control (ATC) in an increasingly capacity-limited air traffic management (ATM) system, air traffic control officers (ATCOs) are progressively supported by automated tools (SESAR Joint Undertaking, 2020). Decades of human-automation research have shown, however, that humans are bad at supervising automated systems and thus benefit greatly from active involvement (Endsley, 2017, Strauch, 2018). On the other hand, routine traffic does take away cognitive capacity from ATCOs that could be better used in handling more complex situations.

One way to redistribute workload and cognitive effort is to allocate a subset of the traffic to an automated agent, enabled by the increased use of controller-pilot data link communications (CPDLC), freeing up the ATCO's cognitive resources, which are needed for complex problem solving. The human ATCO is then responsible for controlling the remaining traffic with active involvement. Exploratory research showed that such a shared airspace is feasible and accepted by ATCOs under certain conditions (Chapter 3). Assigning flights to either a human or an automated agent can be regarded as the next evolutionary phase in flight-centric ATC (FCA), a concept where specific flights are assigned to different human ATCOs (Volf, 2019).

Eurocontrol's Maastricht Upper Area Control Centre (MUAC), an air navigation service provider (ANSP) responsible for the upper airspace of the Netherlands, Belgium, Luxembourg and part of Germany, proposes a strategy to initially only allocate *basic* traffic to an automated system, while the ATCOs are kept engaged with the task of handling the more complex *non-basic* traffic (Hendrickx and Tisza, 2019). Basic flights are presumably easier to automate and do not evoke the creative problem-solving skills that human ATCOs are known to enjoy, making them a prime candidate for delegation to the automation.

The level of responsibility of the envisioned automated system will be increased in three stages, throughout which the automation will only aid with or control the basic part of the traffic. In the first two stages, the ATCO can still take back manual control over a flight. In Stage 1 all flights are handled with approval of the ATCO, in Stage 2 for the basic part there is no ATCO approval, but the controller is still responsible. In the third, and final stage, the automation will autonomously control an entire sector with basic traffic, performing all ATC tasks in that sector. In this stage, the ATCO will no longer be responsible for the traffic and will not be monitoring the sector. An option will be available, however, for the automation to indicate to the ATCO that supervision is required.

As an enabler for this strategy, it is paramount to understand what differentiates a 'basic' from a 'non-basic' flight. Furthermore, this classification should be automated, based on objective criteria that can be obtained in real-time as a flight approaches a sector. Three decades ago, Drew and Makins (1994) already performed an initial study to identify a 'problem-free' set of flights that could be controlled by a planner ATCO through datalink. Current air traffic complexity models, however, predominantly consider the complexity of an entire sector (Hilburn, 2004), for example to predict sector capacity. This is most commonly done by taking a weighted sum of various contributing factors, such as the rate of flights entering/exiting the sector or the traffic density (Lee et al., 2009, Mogford et al., 1995). This sector-wide approach makes them unsuited for classifying individual flights. Studies of per-flight complexity classifications are rare, but suggest their feasibility and call for further research (e.g., Capiot et al., 2019).

This chapter reasons from the perspective of a single flight, rather than an entire sector and asks the following questions: What is the relationship between the number of flights ATCOs consider to impact a single flight and the *perceived* complexity of that flight? What level of consensus exists among ATCOs on these included flights? What traffic parameters impact the perceived complexity the most? To answer these questions, empirically collected flight complexity ratings from fourteen professional en-route ATCOs are analyzed using state-of-the-art supervised learning techniques so as to discover relationships, if they exist, between complexity ratings and traffic factors.

The structure of this chapter is as follows. First, we distill what lessons can be learned from existing complexity measures that are primarily used to describe entire sectors (Section 5.2). Next, in Section 5.3 a human-in-the-loop experiment is described where professional ATCOs had to indicate which other flights they included in their complexity assessment of a single flight of interest (FOI) that varied in location and target state over a number of scenarios. Results of the experiment, and subsequent descriptive performance of our machine learning models are given in Section 5.4. The implications of the findings and an outlook into the future applicability of a flight allocation algorithm are discussed in Section 5.5. Section 5.6 concludes the work.

5.2 Background: modeling flight complexity

5.2.1 From sector-based towards flight-centric complexity

Complexity prediction in ATM has predominantly been done in the context of dynamic sectorization to either split or combine sectors based on expected traffic loads and ATCO workload. Over decades, several complexity models have been developed, such as dynamic density, interval complexity, fractal dimension, input/output approach, Lyapunov exponents and trajectory-based complexity (TBX, Prandini et al., 2011, Prevot and Lee, 2011). The majority of these complexity models output either a scalar value or a map that represent the sector-based complexity by integrating (e.g., counting and averaging) specific flight characteristics over the entire sector, for example Hilburn (2004):

- the number of climbing and/or descending flights
- the variance in heading and speed
- the structure of traffic flows (e.g., crossing angle)
- the number of crossing and/or merge points
- distance at, and time to, the closest point of approach (CPA)

It can be argued that sector-based complexity dilutes the complexity contribution of each individual flight (Isufaj et al., 2022). Take for instance the situation illustrated in Figure 5.1 where the sector-based complexity map indicates a hotspot in the middle of the sector. This, however, does not mean that all flights passing through the center of the sector are equally complex (or, non-basic). Conversely, flights that do not pass through the center are not all basic flights. Additionally, certain sector disruptions, like local adverse weather or an emergency flight, might not impact all flights equally.

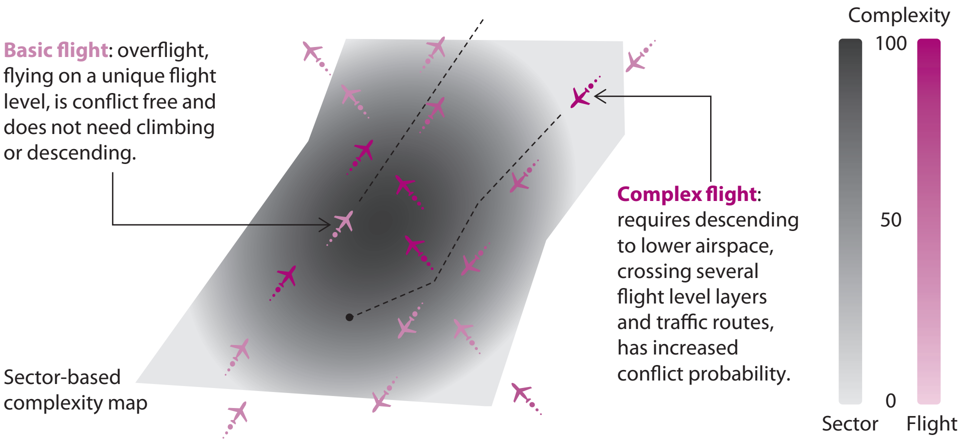


Figure 5.1: Examples of a basic and non-basic (or complex) flight overlaid on a sector-based complexity map.

In an effort to capture the complexity of a single flight, we propose, based on discussions with operational ATCOs and ATM experts at MUAC, that flight complexity centers around *attentional* and *control* demands, such as:

- Attentional demands
  - Trajectory complexity (e.g., winding vs. direct route)
  - Uncertainty (e.g., in climb/descent profiles, arrival time management, pilot delays)
  - Multi-dimensional interaction profile with other flights (e.g., route crossings, altitude overlap, conflict probability)
  - Interaction with environmental disruptions (e.g., restricted airspace, weather cells)
- Control demands
  - Easiness of a conflict resolution (e.g., altitude vs. heading)
  - Conflict geometry (e.g., overtake vs. crossing)
  - Number of required (follow-up) actions (e.g., evade conflict and steer back to target waypoint)
  - Timing of actions (e.g., proactive vs. reactive)
  - Size of the 'solution' space (e.g., sector size for maneuvering flights)

Many of these elements cannot be considered independently in how they impact the complexity of a single FOI and are therefore not easily modeled. For example, given a certain CPA, the convergence angle between flights impacts the time to reach that point. To cope with complexity, ATCOs typically make hierarchical pair-wise comparisons between flights (Rantanen and Nunes, 2005). For example, to detect conflicts, they first scan the flight labels for overlapping altitudes, then narrow down the search to flights with crossing paths, followed by anticipating their CPA (Chapter 4). As such, the ATCOs' strategies, skills and expertise are expected to play a role in how complexity is perceived.

5.2.2 Inherent versus perceived flight complexity

Similar to the division between taskload and the experienced workload (Hilburn, 2004), classification of 'basic' and 'non-basic' flights may depend on the preferences, skills and experience of ATCOs as illustrated in Figure 5.2. This notion suggests that the perceived flight complexity may be individual sensitive, similar to the findings of research that studied the impact of personalization on fostering ATCO agreement and acceptance in the context of conflict resolution advisories (Westin et al., 2016a). Nevertheless, with highly trained professionals, some level of consensus on which flights are more complex than others can still be expected, providing ground for an automated classification algorithm.

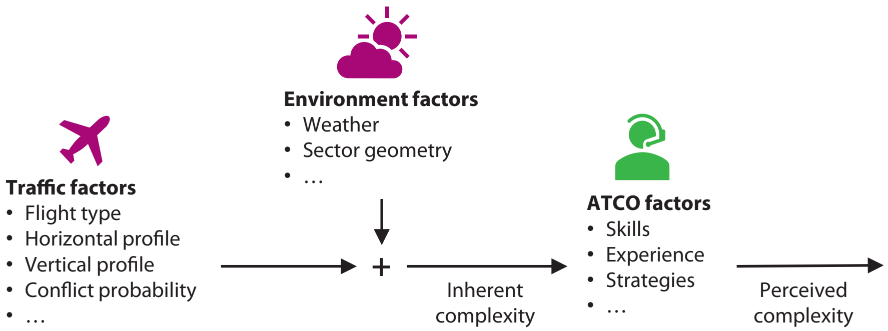


Figure 5.2: Factors that play a role in ATC complexity.

To be able to discern the contribution of inherent and perceived complexity in determining single flight complexity, labeled data would be needed that allows for relating traffic factors to ATCO complexity ratings. Currently, such labeled data does not yet exist. Therefore, the study described in this chapter aimed to collect labeled data on single flight complexity by designing and conducting a human-in-the-loop experiment.

5.2.3 Supervised learning

When labeled data is available, *classification* and *prediction* can be done using supervised learning techniques, such as logistic regression, random forests and gradient boosting trees. These have been used, for example, to determine traffic parameters that are most influential to sector complexity (Pérez-Castán et al., 2022, Pérez Moreno et al., 2022), but to the best of our knowledge not yet on individual flight complexity.

In a classification problem, unbalanced classes can have a detrimental effect on the model's performance. In this chapter, the number of flights not contributing to a single flight's complexity vastly exceeds the number of flights that do matter. When mostly trained on non-relevant flights, a model might not be able to predict the important flights. Using ensemble techniques, such as gradient boosting, the impact of included flights on model training can be increased. Similar techniques are done in medical studies, where decease cases are rare, but of paramount importance to discover and predict.

Note that machine learning is mainly used in this preliminary study to examine and describe the complexity factors in a specifically crafted set of scenarios. Creating an operational prediction model for any traffic sample is outside the current scope and would require more extensive data collection.

5.3 Method

5.3.1 Participants

Fifteen ATCOs from MUAC (Table 5.1) voluntarily participated in a simulator experiment. All participants provided written informed consent and the experiment was approved by the Human Research Ethics Committee of TU Delft under number 2206.

Table 5.1: Participant characteristics.

		Sector group		
	All	Brussels	DECO	Hannover
Number of ATCOs	15	5	5	5
Age, years (SD)	39.9 (6.3)	37.0 (4.3)	40.8 (7.6)	42.0 (5.4)
Experience, years (SD)	15.9 (6.4)	13.0 (4.0)	15.4 (7.2)	19.4 (6.0)

5.3.2 Apparatus

During the experiment MUAC's operational interface was mimicked using SectorX, a medium-fidelity Java-based simulator built by TU Delft (see Appendix A). Figure 5.3 was displayed on a computer monitor and could be controlled with a computer mouse. While only static scenarios were shown, the simulator allowed for some interaction that helped the ATCOs assess the traffic situation. A flight's planned route could be revealed by press-and-hold on the associated label. Furthermore, MUAC's verification and advice tool (VERA) was available to see a prediction of the closest horizontal distance between two flights and their corresponding future positions. And finally, the velocity leaders could be extended to show a flight's predicted position one to eight minutes into the future.

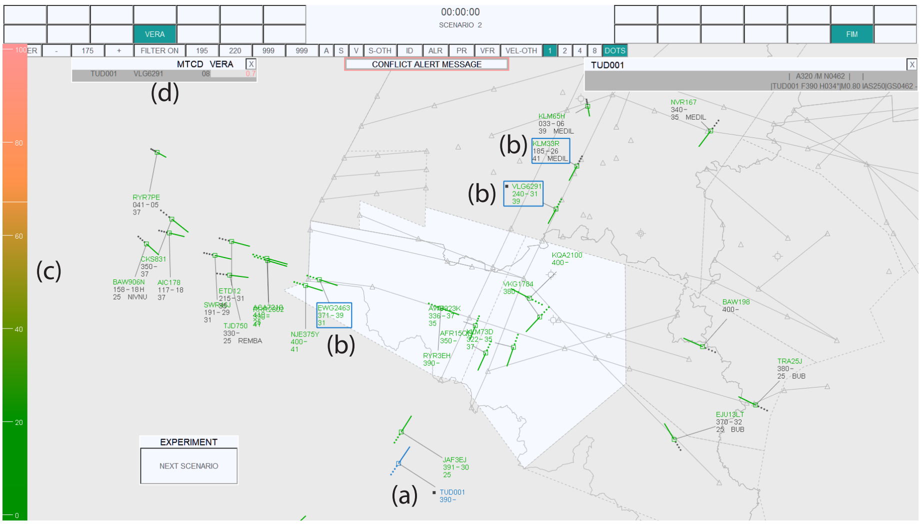


Figure 5.3: Simulator interface showing a Brussels scenario with flight of interest (a), three included flights (b), complexity rating scale (c) and VERA information (d). Background colors have been inverted for print clarity.

5.3.3 Procedure and participant tasks

Each participant followed the same procedure, outlined in Figure 5.4. At the start they were briefed on their task to assess the complexity of guiding an individual FOI from its current location to the required sector exit coordination point (XCOP) and transfer flight level (TFL). The ATCOs then practiced operating the simulator on a simplified scenario containing only two flights in an artificial sector.

Next, four training scenarios were executed, in which the background traffic, sector and experiment procedure were identical to the measurement scenarios. After assessing the situation (optionally making use of the provided support tools), each scenario required two consecutive actions from the ATCOs, before they could click on a button to advance to the next scenario:

1. Indicate which background flights played a role in their complexity assessment (from here on referred to as 'included flights'). If no flights were selected, a confirmation popup was shown before continuing.
2. Register their FOI complexity rating on a 0–100 scale on the screen (see Figure 5.3).

After the four training scenarios were completed, the measurement phase consisted of 36 scenarios. It was followed by a review phase where the scenarios that received the highest, lowest and middlemost rating (three each) were revisited. These nine scenarios were presented in the same order as they appeared in the first phase. The ATCOs could see their registered complexity score and which flights they had included, but were not told why these scenarios were selected for review. The ATCOs were asked to fill out a questionnaire about these scenarios to gain more insight into the reasoning behind the reported complexity and why flights were included or not.

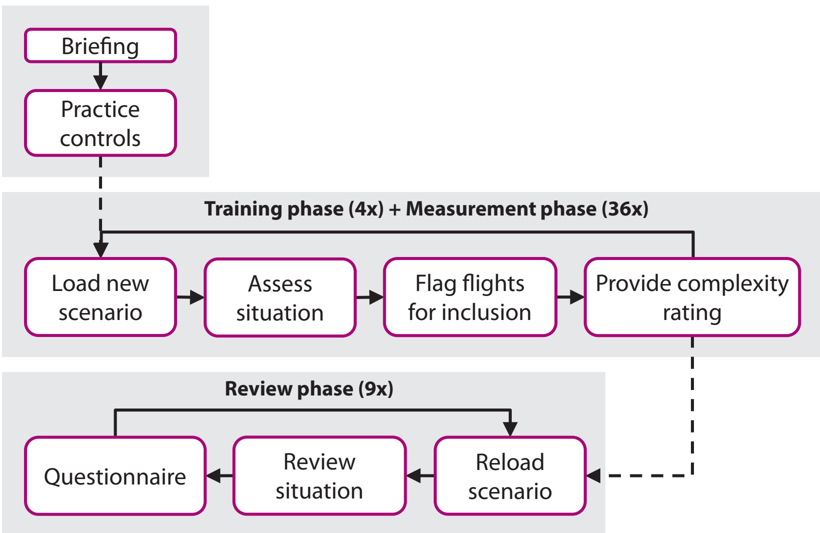


Figure 5.4: Experiment procedure.

5.3.4 Scenario design

Three distinct MUAC sectors were selected for the experiment, ranging from large and relatively quiet (DECO East) to small and dense (Brussels West). The ATCOs were only presented with the sector they had an endorsement for, ensuring comparable familiarity levels. For each sector, a distinct radar snapshot from 23 March 2022 served as background traffic. The snapshots were selected such that it was possible to introduce conflicts with various characteristics. As individual sectors are often combined to balance capacity with demand, we used the same sector configuration as was operational at the time of the corresponding radar snapshot: DECO East contained Jever and Holstein, Brussels West consisted of Koksy and Nicky, and Münster was a single sector from the Hannover sector group. Figure 5.5 shows these sectors and the number of flights in each of them.

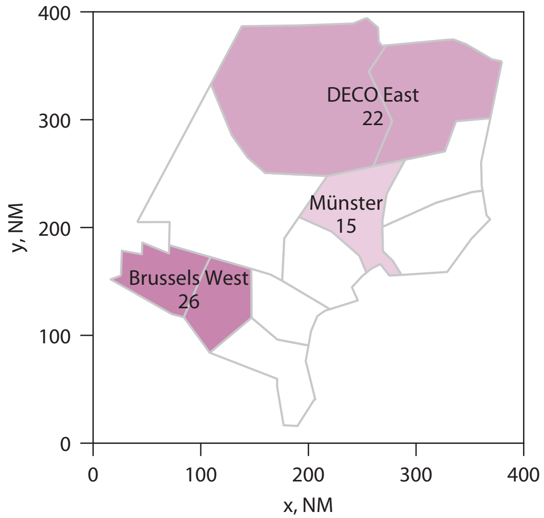


Figure 5.5: Selected MUAC sectors and their number of flights, excluding the FOI.

A single FOI was overlaid on the background traffic in a variety of initial positions and exit conditions to create distinct scenarios (see Figure 5.3 for an example). It was colored differently to distinguish it from other traffic. Manipulating a single flight, instead of using an entirely different traffic sample for each scenario, eliminated the influence of sector complexity factors external to the FOI (e.g., traffic density, total number of climbing flights) as much as possible.

Following the flight complexity demands identified in Section 5.2.1, the various scenarios were manipulated to address all of these demands. For example, by using three different sectors with distinct sizes and traffic densities, control demand in terms of available 'solution' space is manipulated. Sector size also impacts attentional demands since the chance of interactions between flights increases. Within each sector, Table 5.2 lists the traffic factors that were manipulated to impact the complexity of the FOI. Note that these manipulations might have interactions with one another, meaning that it is not possible to study the impact of each manipulation on complexity separately. The scenarios were presented in a partially randomized order to account for order effects.

Table 5.2: Scenario design parameters.

Parameter	Variation	
Time to CPA (urgency)	Short	0–300 s
	Medium	300–600 s
	Long	600–900 s
Conflict-free direct route (easiness of resolution)	Interactions on current trajectory (I)	
	Interactions on current trajectory (II)	
	Interactions on current flight level	
Flight level change (uncertainty)	Small descent	0–4,000 ft
	Small climb	0–4,000 ft
	Large climb or descent	>4,000 ft

5.3.5 Dependent measures

The experiment resulted in the following output measures:

- Complexity rating for the FOI,
- Flights included by the ATCOs as 'contributing to the complexity rating',
- Usage of support tools: VERA, velocity leaders and route preview,
- Questionnaire: reasons to include flights and how comfortable the ATCOs would be to delegate the FOI to the automation.

5.3.6 Data analysis

With the (target) states of all flights known, the features listed in Table 5.3 have been computed to describe each of the flight pairs including the FOI. The selection of features is based on existing sector complexity research referenced in Section 5.2. Lacking sufficient data to accurately predict climb or descend points, horizontal positions are extrapolated along the current tracks and ground speeds. No advanced trajectory predictions are used yet in this exploratory study. To exclude predicted conflicts beyond a reasonable look-ahead horizon, the calculation of the features was limited to the trajectory before reaching the XCOP for flights descending to a lower airspace within the sector. If a predicted CPA would occur after reaching the XCOP, the CPA was capped to the distance between the flights upon reaching the XCOP. This was only done for these descending flights, as ATCOs do 'look' beyond their sector boundaries to prevent causing any conflicts for their colleagues in adjacent sectors.

Table 5.3: Candidate features of included flights relative to the FOI.

	Feature	Unit	Comment
Attentional demands	Current horiz. separation	NM	
	Predicted min. horiz. separation (CPA)	NM	
	Time to CPA	s	
	Vertical separation	ft	
	Exit altitude difference	ft	
	Overlapping flight levels	true/false	True if flights *may* be at the same level at some point
	Climbing/descending	true/false	
Control demands	Convergence angle	deg	
	Ground speed difference	kts	
	Flight state	-	Assumed or transferred to me
	Distance to XCOP	NM	Along a direct path
	Required altitude change	ft	From actual flight level to TFL

To relate a single FOI complexity rating to the characteristics of multiple included flights in a scenario, the aggregated features listed in Table 5.4 have been proposed. Note that these only relate to the FOI itself, or in relation to flights included by the ATCO. Non-included flights may have an impact on the sector complexity, but have been considered irrelevant to the FOI complexity in this study. For all altitude differences the absolute value was taken.

Table 5.4: Candidate features for the FOI, aggregated over all included flights.

	Feature
FOI	Required altitude change
	Distance to XCOP
Included flights	Number of flights with altitude overlap
	Number of climbing flights
	Number of descending flights
	Number of flights with CPA under 10 NM
	Number of flights with identical TFL
	Min./average current separation
	Min./average CPA
	Min./average distance to XCOP

5.4 Results

The results are discussed in three steps:

1. The experimentally *collected data* is described in terms of number of included flights with respect to the complexity rating and the level of consensus between different ATCOs, in addition to their use of support tools.
2. A *classification model* is used to determine whether the inclusion of a flight can be linked to objective features and what the relative importance of each feature is.
3. In combination with the FOI's complexity rating, a *regression model* is used to examine the feasibility of predicting a FOI's complexity through its included flights.

Since the number of flights varied over the sectors (see Figure 5.5), percentages of the total number of flights shown to the ATCOs are used when comparing sectors. Each cell in the tables refers to values belonging to one ATCO, unless explicitly stated otherwise.

5.4.1 Complexity rating and number of included flights

Figure 5.6 shows the subjective complexity ratings as given by the ATCOs for each of the 36 scenarios of their sector that were presented to them. The large spread in ratings per ATCO shows that the designed FOI manipulations had an effect on the perceived complexity, as the background traffic did not change between scenarios. This effect was weaker in Münster, where the ATCOs gave relatively low ratings compared to their colleagues in the other sectors. To account for between-participant differences, complexity ratings per ATCO are standardized by z-scores in the remaining analyses.

In total, the ATCOs included 1,139 (10.0%) of the 11,340 flights that were presented to them. Although the number of flights was different for each sector, the share of included flights seems to be primarily ATCO-dependent and varies as much as between 6.8-19.2% in a single sector (Table 5.5). One Brussels ATCO is a noticeable outlier with 180 included flights, significantly skewing the average for that sector.

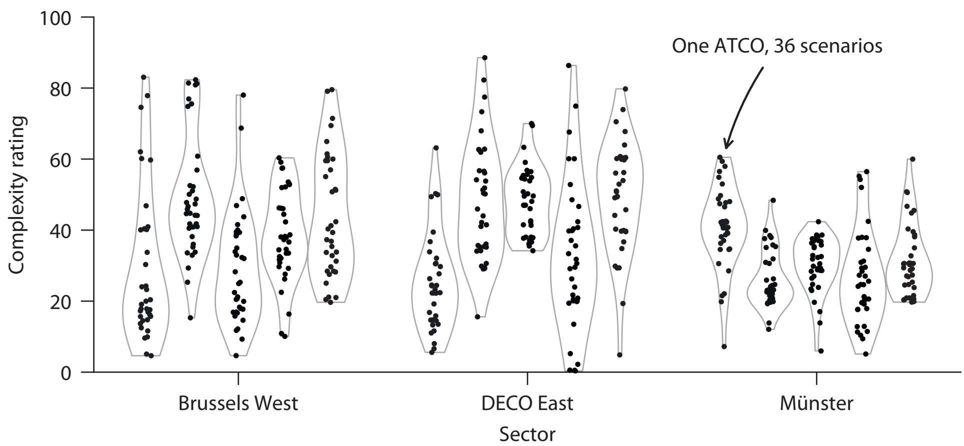


Figure 5.6: Subjective complexity ratings per ATCO for each scenario.

Table 5.5: Total included flights per ATCO, as share of total flights presented to that ATCO.

	Brussels West		DECO East		Münster	
	118	12.6%	60	7.6%	55	10.2%
	93	9.9%	70	8.8%	41	7.6%
	180	19.2%	97	12.2%	37	6.9%
	64	6.8%	56	7.1%	52	9.6%
	71	7.6%	77	9.7%	68	12.6%
Mean	106	11.2%	72	9.1%	51	9.4%
Std. dev.	42	4.5%	15	1.8%	11	2.0%

Similar to the complexity ratings, the number of included flights can be standardized per ATCO to account for individual differences. A Kendall's tau-b correlation test, chosen because of the non-normality of the data, then shows a moderate positive correlation ($\tau_b = .547, p < .001$) between the complexity scores given by the ATCOs and the number of included flights for all sectors combined. In Table 5.6 the correlations are given per ATCO and sector. The strongest, yet still moderate, correlation is found for Brussels West ($\tau_b = .611$). The correlations are statistically significant ($p < .001$) for all sectors. In both DECO East and Münster, one ATCO exhibits noticeably weaker correlations than the other ATCOs.

Table 5.6: Correlation between standardized number of included flights and standardized complexity score per ATCO.

	τ_b, p		
	Brussels West	DECO East	Münster
	.592, <.001	.627, <.001	.381, =.004
	.684, <.001	.802, <.001	.561, <.001
	.702, <.001	.478, <.001	.563, <.001
	.718, <.001	.388, =.003	.553, <.001
	.592, <.001	.547, <.001	.773, <.001
All ATCOs	.611, <.001	.503, <.001	.527, <.001

5.4.2 **Usage of support tools**

To determine (or confirm) whether a flight conflicts with the FOI, the ATCOs could use VERA to show the predicted minimum separation between two flights. Usage varied greatly over the ATCOs, ranging from not being used at all to checking 106 flight pairs (Table 5.7). Note that the sectors cannot be readily compared with each other, due to their vastly different number of flights (and thus potential conflicts). The Brussels ATCO, who included the most flights, was also by far the most active user of VERA. All ATCOs practiced with VERA in the training phase and were thus aware of its availability. In total, 352 (69%) of the 512 VERA flights were eventually included (Figure 5.7), while only 8% of the flights with a CPA below 10 NM was *not* included by the ATCO after confirming this distance through VERA. Presumably, some ATCOs felt comfortable with smaller separation margins and/or would not consider this an immediate problem for far-away flights. Above 20 NM, only some flights were included, mostly by ATCOs who considered any VERA-check to be a part of their cognitive complexity assessment.

Table 5.7: Number of flight pairs on which VERA was used per ATCO and how many of those were included.

Checked pairs (included)		
Brussels West	DECO East	Münster
38 (37, 97%)	11 (10, 91%)	9 (8, 89%)
24 (14, 58%)	35 (29, 83%)	0 (0)
106 (61, 58%)	1 (0)	33 (25, 76%)
63 (41, 65%)	56 (36, 64%)	9 (8, 89%)
78 (47, 60%)	49 (36, 74%)	0 (0)

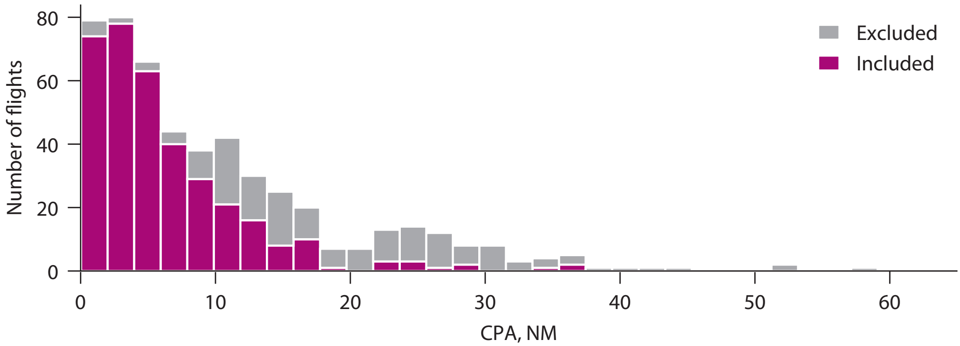


Figure 5.7: Number of flights on which VERA was used and how many of those were subsequently included, with respect to their minimal separation distance to the FOI.

Besides VERA, extending the velocity leaders beyond the default one minute is another, more crude, technique to check future positions of flights and assess their CPA. As the velocity leaders are adjusted for all flights at once, this cannot be linked to the inclusion of particular flights. Neither did we find indications for velocity leaders being used instead of VERA (i.e., a DECO ATCO who used VERA only once did not extend the velocity leaders at all).

Akin to the usage of VERA, the number of flights for which a visual representation of the planned route on the radar display was requested varied considerably between 0 and 54. While the display of routes made flights with planned turns more pronounced, the planned turn was already indirectly visible by the waypoint listed in the flight's label. To illustrate the limited predictive value of this measure: three ATCOs requested the route of the same flight that would come into proximity once the FOI commenced a turn, but only one of them decided to include it.

5.4.3 **ATCO consensus**

As the number of included flights already shows, there is a level of subjectivity in the data. We therefore introduce three majority levels regarding flight inclusion. Consensus is reached when all five ATCOs of a sector unanimously agreed to either include or exclude a flight. For a qualified or simple majority respectively four and three ATCOs were in agreement. The distributions in Figure 5.8 show a high level of consensus for all sectors, with the ATCOs unanimously agreeing for 84-88% of the flights, increasing to 94-96% with qualified majorities. Between any two ATCOs in a sector, 88-97% of the flights were identically labeled. The relatively low share of excluded flights in Brussels West, compared to the other sectors, is mostly due to the large number of inclusions by a single ATCO, as also reflected in Table 5.5.

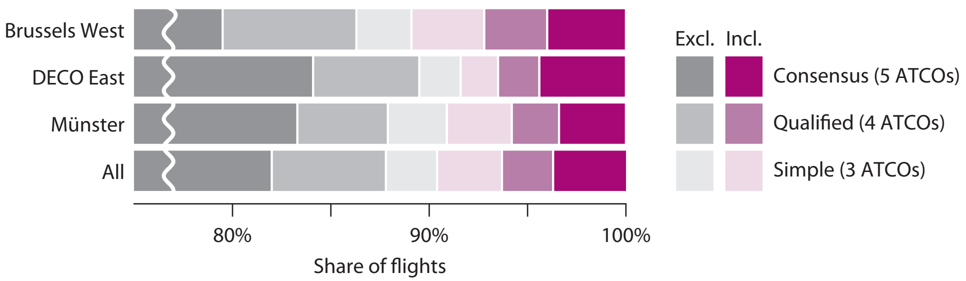


Figure 5.8: ATCO consensus on flight inclusion per sector and for all sectors combined.

Figure 5.9 shows the number of flights per scenario that was included by a qualified majority of the ATCOs (i.e., four or five ATCOs), versus the average standardized complexity rating for that scenario. Note that the ATCOs agreed on the inclusion of just a single flight in the vast majority of scenarios. Again, a moderate positive correlation is visible ($\tau_b = .553, p < .001$), but it is also clear that a higher number of included flights does not necessarily relate to a higher complexity rating. Hence, the features of those particular flights might better explain some of the variability.

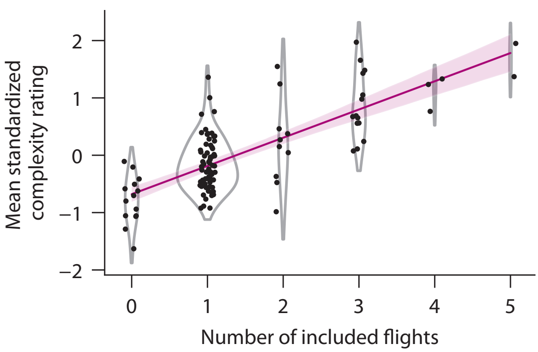


Figure 5.9: Correlation between standardized complexity rating and number of flights included by a qualified ATCO majority in each scenario. Shown with 95% confidence interval.

5.4.4 Features of included flights

To analyze which features play a role in the ATCOs' selection of included flights and see whether this selection can be modeled, we applied a gradient boosting classifier on the features from Table 5.3. Gradient boosting was used in this study, because of its ability to combine weak learners (e.g., due to imbalanced data) into a strong model. The model target was to classify whether a flight was included or excluded by a specified majority of the ATCOs. All flights that did not meet the specified level of consensus were filtered out to ensure that the model was trained and tested on a progressively well-labeled data set. As the label was binary (include or exclude), the simple majority case included all flights.

To avoid under- or overfitting the model, the data was split over four stratified folds, meaning that the share of included flights was equal in all folds. The model was then trained and tested on four splits (each consisting of three training folds and one testing fold) and subsequently tuned through cross-validation and grid search for high F1-scores (a balance between precision and recall).

The resulting confusion matrices, summed over the four splits, are shown in Figure 5.10 for each of the majority categories. This clearly reflects the expected increase in performance when filtering on at least a qualified majority that provides more robust labels on the data (Table 5.8, averaged over the four splits). 89% of the flights that were included by all ATCOs were correctly classified as 'include' by the consensus model, while only 11% of the included flights were missed.

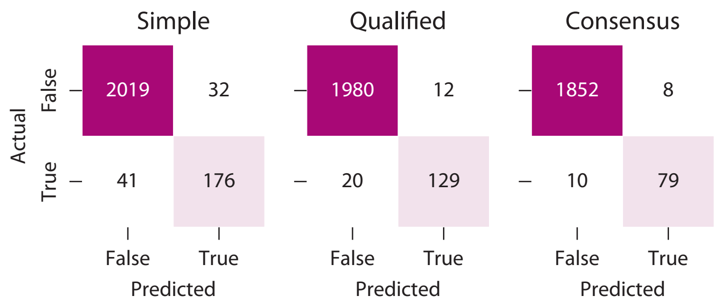


Figure 5.10: Classifier confusion matrices per majority category.

Table 5.8: Flight inclusion classifier performance.

Majority category	Accuracy	Precision	Recall	F1
Simple	0.97	0.86	0.81	0.83
Qualified	0.99	0.91	0.87	0.89
Consensus	0.99	0.91	0.89	0.90

As a measure for the predictive value of each of the features, their relative importance in the consensus model is given in Figure 5.11, as an interval over the four folds. As was expected, the predicted minimum separation (CPA) appears to be the most important feature, followed by the presence of an altitude overlap. Flights where the altitude bands are not overlapping will never be in conflict, unless one of the flights has to deviate

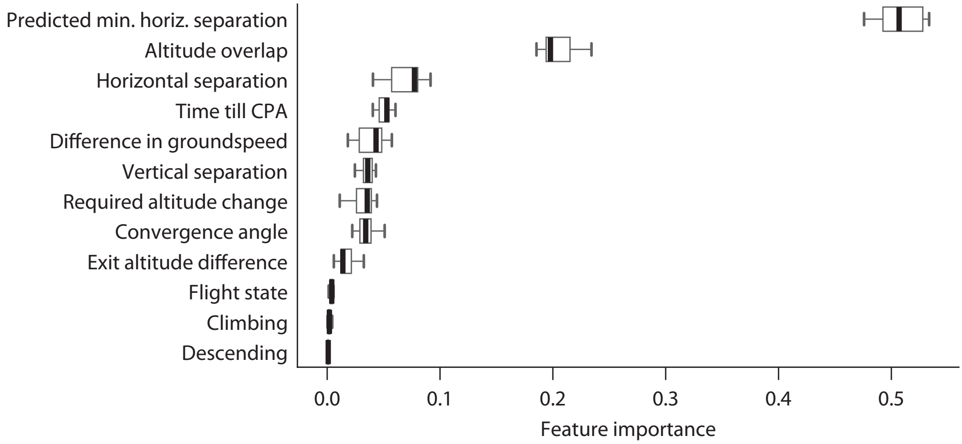


Figure 5.11: Feature importance of consensus classifier model.

to another level. To illustrate, only seventeen (0.4%) out of 4,492 flights without altitude overlap have been included by the ATCOs and never by more than one ATCO at a time. Ten of these were in the Jever scenarios, with a single ATCO including seven. We were unable to identify the reasons for including these particular flights, other than two cases where the included flight would first climb and then descend within the controlled sector, whereas our metric purely looked at current, cleared and exit flight levels to assess the overlap. The current horizontal separation and time till CPA are marginally more important than the other features used in this study. Finally, a flight's ATC state and whether it is climbing or descending seem to have negligible impact.

5.4.5 **Predicting complexity ratings**

Besides identifying the important features of flights that may impact whether a certain flight should be included or not in assessing the FOI complexity, it would also be important to predict the complexity rating associated with the FOI. In that way, the future system envisioned by MUAC would be able to predict the complexity level of a flight entering the sector, classify it as either 'basic' or 'non-basic' and assign the flight to either the automation or the human ATCO, respectively.

Creating such a prediction model first requires that parameters denoting the relationships between the FOI and the included flights are aggregated by descriptive statistics, such as the average, sum, minimum, etc. Table 5.4 lists the relational parameters that we included in this first exploratory study. When an ATCO included zero flights for a scenario, it was filtered out in this study, as no aggregated features could be computed in that case. The model's goal is mainly to detect the complex cases and scenarios with zero included flights received relatively low complexity ratings anyway. This model is, furthermore, independent of the level of consensus between ATCOs, as they may agree on the inclusion of some flights in a scenario, but may also include flights in their complexity rating for which no consensus was reached. Therefore, we consider their individual combination of included flights and complexity rating.

To test and train the gradient boosting regression model, a fifteen-fold is used with one ATCO per fold. This ensures that the data belonging to a single ATCO does not get spread out over the training and test data, such that the model performance is an indication for how well the model generalizes on ratings from other ATCOs for which it was not trained. The model's hyperparameters were tuned through cross-validation and grid search optimizing for high R^2-scores.

Figure 5.12 shows an example of the actual complexity ratings by the ATCOs and the corresponding model-predicted rating, for each of the folds combined. With $R^2 = 0.16$, $MSE = 0.68$, $MAE = 0.65$ and $RMSE = 0.34$, the model's performance is relatively weak compared to existing subjective sector-based complexity models, such as those discussed in Andraši et al. (2019). The importance of the features, over the fifteen splits, is given in Figure 5.13. Despite the weak model performance, some observations stand out. First, the number of flights with an altitude overlap is clearly the most important feature. According to Section 5.4.4 this is closely related to the number of included flights, confirming this metric's moderate correlation with the complexity rating. Furthermore, flights at a closer distance to their exit point (XCOP) are more likely to receive a high complexity rating. Presumably because their solution space is limited. Finally, a group of features is of equal or marginally different importance, confirming that many factors play a role in perceived complexity.

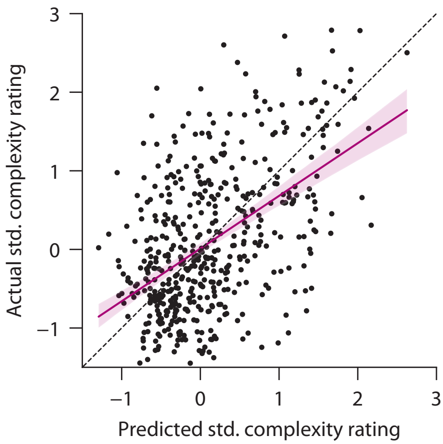


Figure 5.12: Comparison between the original data and regression model output, with 95% confidence interval.

5.4.6 Willingness to delegate flights to automation

With the complexity rating known, a flight can be classified as either basic or non-basic based on a complexity threshold. In the post-measurement reviews, the ATCOs had to indicate how comfortable they were with having the FOI handled by the automation. Unfortunately, due to technical issues, only a small part of this data was saved. Based on this limited data and discussions with ATCOs, a higher complexity rating seems to generally match with a lower willingness to delegate the flight, tipping around the zero in their z-scored ratings.

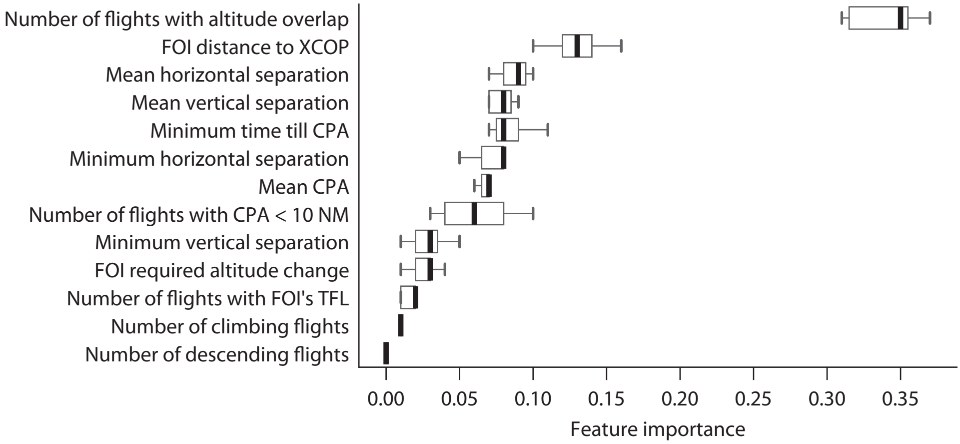


Figure 5.13: Feature importance of regression model.

5.5 Discussion and recommendations

5.5.1 Included flights

Despite a high level of consensus, the ATCOs clearly had different interpretations of which flights to include. This can originate in different working styles, with some ATCOs more proactively solving distant conflicts, but may also indicate a lapse in the briefing. Several ATCOs, for example, included all flights that would lead to a loss of separation if no action was undertaken, while other ATCOs did not include such flights if a straightforward solution was available (e.g., descending the flight to its TFL). Furthermore, some ATCOs included flights that did not directly pose a problem for the FOI, but that decreased the solution space for solving conflicts between the FOI and other flights. This was especially evident when the FOI had to fly opposite a stream of bunched flights.

The features that we selected proved sufficient to classify most of the flights for which there was consensus between the ATCOs though. The relative importance of the CPA and the presence of an altitude overlap that was found in the included flight analysis, strengthens the hierarchical task analysis presented in Chapter 4. ATCOs seem to predominantly filter flight pairs based on these two characteristics. Nevertheless, we identified several possible improvements during our analysis. Most prominently, we simplified conflict prediction to a mere extrapolation along the current track, ignoring any expected turns or speed changes that the ATCOs may have included.

5.5.2 Complexity ratings

The results show a moderate correlation between the number of included flights and complexity ratings. This confirms the idea behind the Dynamic Density model, where number of flights in a sector is the primary driver for complexity (Prandini et al., 2011). Brussels West showed the strongest correlation, which is most likely related to its relatively large number of flights, and therefore interactions, compared to the other sectors.

Again, a discrepancy in the used definition of ‘complexity’ cannot be ruled out. Although standardizing the ratings per ATCO is an established method to reduce between-participant differences, it cannot ensure that all ATCOs equally isolated the complexity of the FOI from that of the entire sector.

The gradient boosting model was able to predict the complexity ratings to some extent, but would have to be improved if it were to be used for flight classification. The input features need to include additional measures for both the FOI and other flights. For example, whether a flight can transit to its TFL unhindered, or whether it has to be put on a heading, requiring prolonged monitoring and rejoining the route. These factors are known to add to the perceived complexity (Fothergill and Neal, 2013).

5.5.3 Experiment design

The present study only considered a single base traffic sample per sector and is therefore not necessarily applicable to every traffic situation within or outside these sectors. The fact that we observed differences between the three sectors can stem from multiple factors, including the participating ATCOs, the sector geometry or the used traffic sample.

While the base traffic was a snapshot from real traffic data, the artificially introduced FOI did not always match ATCO expectations. Some scenarios were rated more complex than initially expected, because they presented an abnormal situation to the ATCOs. In a small number of scenarios, the FOI was planned to fly a non-straight trajectory. While the retrieval of routes was logged in the experiment, the route points could also be seen in the label without any (logged) action. If ATCOs incorrectly assumed the flight would proceed along its current track, it would most likely have affected their choice of inclusion, as some conflicts only existed along the planned trajectory. Since we primarily focused our analysis on flights for which there was consensus, we expect its impact to be limited, however. Nevertheless, future research should aim to only include realistic FOIs to completely eliminate such inconsistency. For example, by taking a large sample of radar snapshots and highlighting a single FOI coming towards or just entering the sector.

The relatively small number of participants per sector increased the potential influence of outliers. With a larger sample size, the qualified majority may be more usable. This would increase the certainty about which flights to include beyond just the unanimously included flights.

5.5.4 Operational relevance

As soon as we can predict an individual flight's complexity based on objective, readily available traffic characteristics, the next step would be to determine the threshold, below which flights are considered basic. The incomplete questionnaire data from the reviewed scenarios does not provide sufficient ground for this cause, other than the observation that the willingness to delegate flights was largest with low complexity. This finding is consistent with the results of Chapter 3 and matches MUAC's proposed strategy to automate basic flights first (Hendrickx and Tisza, 2019). The ATCOs indicated that, among others, high trajectory uncertainty of potentially interfering flights was a key reason to be hesitant about delegating a flight. The introduction of automation-directed flights within an airspace may itself have an impact on the perceived complexity of human-directed flights due to the changed teamwork dynamics and tasks (Prandini et al., 2011) and associated uncertainty. This effect is not included in our current analysis and strongly depends on the way the automated system is implemented.

In an operational context, it would make sense to automatically assign basic flights to the automation, while leaving non-basic and undetermined flights to the human ATCO. Manually handling a basic flight is expected to be a smaller nuisance than prematurely allocating a non-basic flight to automation. Thus, the model should be tuned favoring a high true positive rate (i.e., recall metric of a classifier) over a high precision. Here, expert opinions play an essential role in establishing the threshold to increase ATCO acceptance.

Finally, tweaking the model to the individual ATCO might result in a more accurate model and hence increased ATCO acceptance (Westin et al., 2016a). On the downside, a personalized model might create an unworkable situation where flight allocations change whenever a new ATCO takes over from a colleague. It also means that the automation has to be sufficiently advanced to handle a wider range of complexities than when it is limited to flights about which consensus was reached.

5.6 Conclusions

In the development of a future ATC system where human controllers remain in charge of all non-basic flights while the automation handles all basic flights, this chapter demonstrated the feasibility of classifying basic and non-basic flights, based on features extracted from their interaction with surrounding traffic. We showed that the perceived complexity of a single flight of interest can be related to the combined sum of interactions that this flight has with other traffic.

Follow-up research should determine the complexity threshold below which flights can be considered basic. Subsequently, the operational applicability should be validated by simulating a shared human-automation airspace with flights automatically assigned to either agent based on the presented model. With increasing model accuracy leading to a larger share of confidently classified flights, increasingly more flights can be automatically allocated to the automation.

6

Validation

After the exploratory research in Chapter 3 and the more focused and controlled studies from Chapters 4 and 5, this chapter discusses the most realistic and most comprehensive experiment of the thesis. In this final experiment, 14 air traffic control officers are subjected to two allocation schemes: one pragmatic, based on flows (i.e., overflights and inbound/outbound streams), and one carefully constructed to minimize interactions between automated and manual flights. The chapter serves to validate the findings of all other chapters regarding flight allocation best-practices, and to assess the practical use of an interaction-based allocation scheme in future air traffic control operations.

6.1 Introduction

En-route air traffic control (ATC) is currently primarily sector-based, with all flights within a geographic area under the responsibility of a single *executive* air traffic control officer (ATCO), often supported by a second *coordinating* ATCO who mainly coordinates with adjacent sectors (Pfeiffer et al., 2015). With the projected growth in air traffic over the next decades and a worldwide shortage of ATCOs, this sector-based flight allocation is under pressure (Pavlović et al., 2023). Flight-centric operation, where one executive ATCO guides a flight from departure till arrival, aims to improve the workload balance between available ATCOs (Finck et al., 2023b, Schmitt et al., 2011). Advances in computer capabilities and an increased use of pilot-controller data link communications (CPDLC), however, pave the way for the introduction of higher levels of automation (LOAs) in ATC.

In future ATC systems, part of the traffic may therefore be completely delegated from the human ATCO to an automated computer system to reduce ATCO workload (Hendrickx and Tisza, 2019). The resulting mixture of authority over flights within a single airspace, however, introduces new and potentially more intricate interactions. Non-complex flights, such as high-altitude overflights, are the prime candidate for delegation, as they are relatively easy to automate, require little input from ATCOs and have generally few interactions with other flights in the airspace (Chapter 5). Leaving more complex traffic directed by the ATCO also helps them to remain proficient, alert, and increases acceptability of automation.

An empirical workflow analysis of the conflict detection and resolution tasks (Chapter 4) supports the notion that mixed conflicts, where flights with interacting flight trajectories are allocated to different agents, are particularly demanding for ATCOs. Together with the associated lower predictability of vertical flight paths, it is hypothesized that it is undesirable to split the responsibility over flights involved in such interactions between human and automation. Providing a single agent (human *or* automation) with full control over interacting flights allows the agent to potentially prevent conflicts (Martins et al., 2019) or, if a conflict does occur, instruct the flight that has the most efficient or easy-to-monitor solution. While various works have touched upon the area of delegating specific flights to automation (Vanderhaegen et al., 1994) or another ATCO (Finck et al., 2023b), no extensive previous work is known that empirically tests and compares human-automation flight allocations in a realistic current-day ATC setting.

This chapter presents two flight allocation schemes (Section 6.3). One is based on the aforementioned insight that flights should be delegated based on their level of complexity, which is mainly driven by interactions with other flights (Chapter 5). The other is a more pragmatic scheme based on the notion that overflights are generally the least complex, together with clearly structured and therefore predictable inbound and outbound streams. Through a simulator experiment with fourteen professional en-route ATCOs, as described in Section 6.4, the potential performance and ATCO acceptance benefits of the interaction-based allocation is investigated in Section 6.5. Unlike the experiment from Chapter 3, the human/automation allocation was fixed per scenario and could not be adjusted by the ATCO (Figure 6.1). Section 6.6 discusses the impact of the results and limitations from the experiment, with the final conclusions and outlook for future use given in Section 6.7.

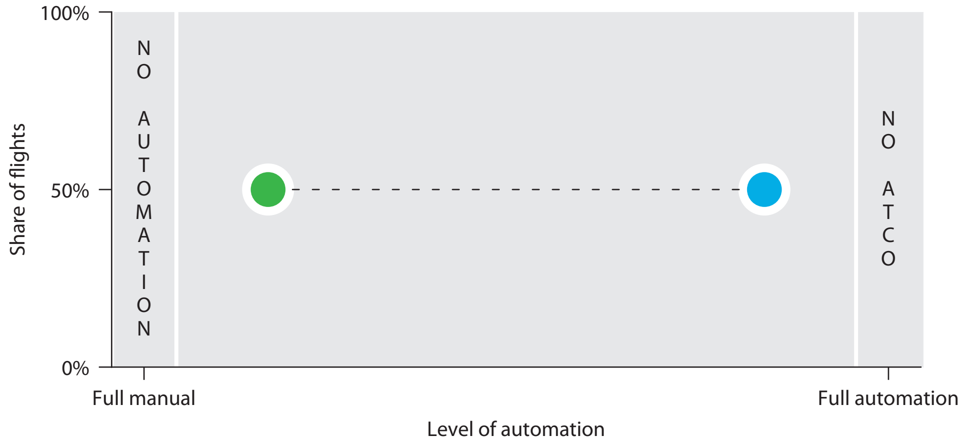


Figure 6.1: Experiment levels of automation, as introduced in Section 2.3. The ATCOs could not adjust the initial allocation of flights, which was designed for an approximately equal distribution between both agents.

6.2 Background: Brussels sector group

The Brussels sector group from Eurocontrol's Maastricht Upper Area Control Centre (MUAC) serves as an example. This sector group covers the airspace above 24,500 ft over both Belgium and Luxembourg and is relatively small and dense compared to the other MUAC sectors. In our previous work on single flight complexity, Chapter 5, this group showed the highest correlation between flight characteristics and their perceived complexity, paving the way for developing an algorithm that can automatically delegate specific flights to automation.

The group can be split into four sectors: Koksy, Nicky, Olno and Luxembourg, which all can be further split into a low (FL245–355) and high (FL355+) sector. Based on traffic demand, one executive ATCO can control a combination of multiple sectors in various configurations, e.g., *Brussels West*, comprised of Koksy and Nicky. Shown in Figure 6.2, key traffic characteristics of Brussels are a large east/west stream coming from and going to the London airports and a considerate north-south stream mainly consisting of Amsterdam and Paris in- and outbounds. Traffic to/from Brussels and Düsseldorf further complicates the sector. Traffic density is highest in the western part of the sector (Koksy and Nicky). Flights coming from the west are generally still climbing as they enter the airspace, while eastbound flights are mostly cruising flights that will not start their descent into London before they have been transferred by Brussels shortly after crossing the border between Nicky and Koksy.

6.3 Flight allocation schemes

A good allocation scheme results in a fair share of flights allocated to either a human or automated agent. Too few automated flights means that the added value of automating them becomes small, while too many automated flights can make human ATCOs complacent, bored and lead to skill-erosion. In all cases, the success of the allocation depends on the capabilities of the automated agent. If the automation is relatively basic, a more

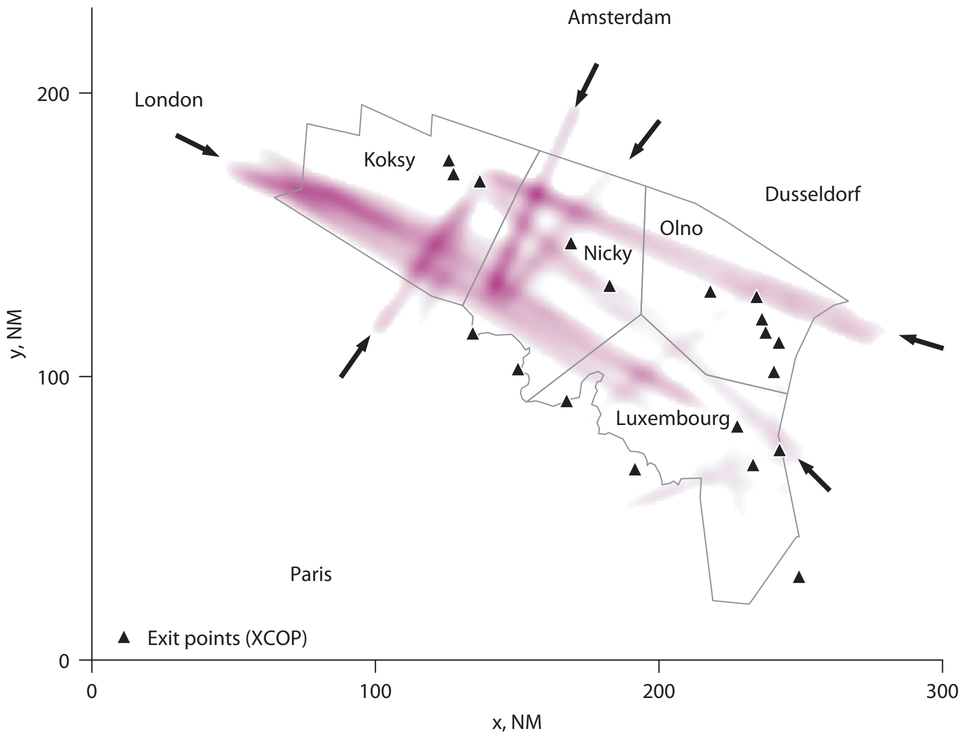


Figure 6.2: MUAC's Brussels sector group with main traffic flow densities of flights assumed by its ATCOs on Saturday 2 September 2023 between 17:15–17:40 UTC.

segregated allocation can be beneficial to prevent mixed conflicts where the automation has a different plan than the ATCO. When two flights interact, it is easier and more efficient to have both flights under one controlling agent, as this agent can then choose which flight to adjust (or even both) to most efficiently solve the conflict.

This problem is also relevant in flight-centric operations (Martins et al., 2019), where a flight is assigned to a single ATCO over its entire route to balance the workload between available ATCOs. If two flights, handled by two different ATCOs, are in conflict, a 'less impacted flight algorithm' will propose which ATCO can best solve the conflict and will inform the other ATCO to observe (Finck et al., 2022). Similar heuristics could be used when either of the ATCOs is an automated system, although the capabilities of the system will be an additional factor in determining who could best solve the conflict.

In this chapter, two static allocation schemes are investigated: 1) a *flow*-based allocation, based on overflights and clearly defined traffic streams, and 2) an *interaction*-based scheme, that takes into account the predicted interactions between flights, building upon previous complexity-based work in Chapter 5. The static nature of the allocation means that flights do not get re-allocated after entering the sector. The two schemes discussed here are used as independent variable in the experiment described in Section 6.4. Since there is no ubiquitous research on what level of flights can best be automated in such a setting, a 50-50% division has been chosen as the starting point.

6.3.1 Flow-based allocation scheme

When asked about which flights to automate first, the majority of ATCOs mention overflights as being prime candidates, i.e., flights requiring limited level changes within the controlled sector (Chapter 3). There is, however, no consensus on what this limit should be. Given the aforementioned traffic sample provided by MUAC (Figure 6.2), during 25 minutes, only 12 flights (14%) passing through the Brussels airspace had no planned flight level change ($|\Delta FL| = 0$ ft). Automating such a small fraction of flights can be a first step towards higher levels of automation, but will bring a minimal reduction in ATCO workload, as these flights already require hardly any intervention, if at all. The $|\Delta FL|$ threshold can be progressively increased to classify a larger share of flights as overflights. Anecdotal evidence from ATCOs suggests that any vertical change between 2,000 and 5,000 ft can still be considered as overflights (Chapter 3). Raising the limit to 5,000 ft yields 32 overflights (36%) in the given traffic sample.

As this is still well below the 50% target distribution, a further increase can be obtained by also allocating clearly defined streams of traffic to the automation. Gil et al. (2023) found that allocating flights to ATCOs solely based on their entry heading was not positively received by ATCOs. In flow-centric operations, flights following similar trajectories are assigned to the same ATCO which, compared to flight-centric operations, leads to fewer mixed conflicts (Bin Jumad et al., 2023). Adding inbounds to and outbounds from Amsterdam (EHAM), for example, raises the share of automated flights to 51% for the given traffic sample.

6

Figure 6.3 illustrates this flow-based allocation concept. By adjusting the overflight flight level change threshold and/or adding or removing traffic streams, the share of automated flights could be controlled. This can be sector- or ATCO-dependent, but may also change dynamically in response to inbound/outbound peaks at airports, similar to the long-standing practice of splitting or collapsing (i.e., merging) sectors when the traffic demand changes. Because the allocation of each individual flight does not depend on the other flights in this scheme, the allocation can be determined whenever a flight approaches the sector.

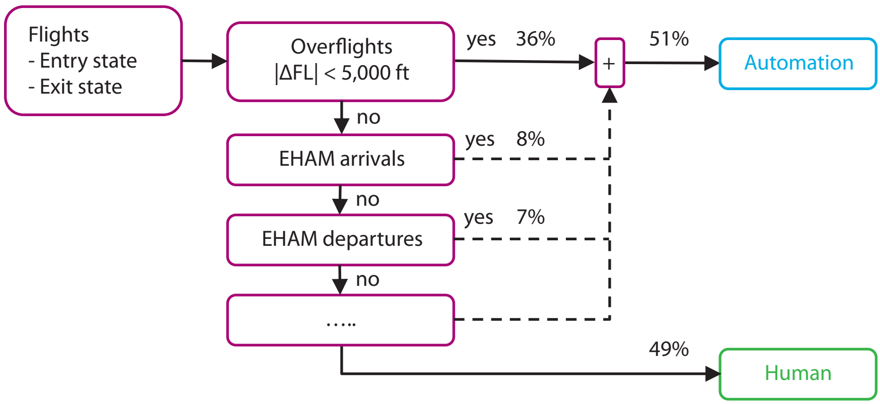


Figure 6.3: Flow-based allocation scheme with corresponding traffic distributions for the scenario described in Section 6.2.

6.3.2 Interaction-based allocation scheme

Although the flow-based scheme should, thanks to rigorous airspace design, segregate traffic to some extent, conflicts can and will still occur. This especially applies to flights from different directions and at different speeds that need to be merged in a single stream, such as the many flights coming from the direction of London. Ideally the complexity of flights is predicted real-time based on the study from Chapter 5, but for this chapter a fast-time simulation was used to imitate such an algorithm and determine in advance which flights would interact based on their flight plans.

For this, the expected trajectories of all flights are simulated, ignoring any conflicts that may arise. Loosely mimicking ATCO best-practices, all flights are cleared to climb to their expected cruise level as early as possible and descend as late as possible, to eventually arrive at the planned transfer flight level (TFL). In the lateral plane, two distinct strategies are used: 1) flights follow direct routes to their exit coordination point (XCOP) as soon as this can be done without clipping the sector boundary, and 2) flights follow their filed flight plan route for the entirety of the sector. This mimics a strategy where ATCOs will try to send a flight direct, but if that is not possible, or their attention is elsewhere, the flight will simply continue along the filed route.

Using short term conflict alert (STCA) predictions from Eurocontrol (2007), the 'overlap' between trajectories of each combination of flight pairs is then determined. As not only flights breaching the regulatory separation minima are considered to be interacting (see Chapter 5), the minimum separation distance is raised from 5 to 10 NM and the look-ahead time is increased from 2 to 8 minutes (in line with MUAC's lateral obstacle display Eurocontrol, 2024b), compared to the standard STCA configuration. Flight *pairs* that fall within this detection window at some point along one of the two simulated trajectories are assigned to either the automation or the ATCO, such that roughly 50% of all the flights belong to either agent. Figure 6.4 illustrates this interaction-based allocation concept. Similar to the flow-based allocation scheme from Section 6.3.1, the threshold (and STCA configuration) can be changed to adjust the distribution between both agents.

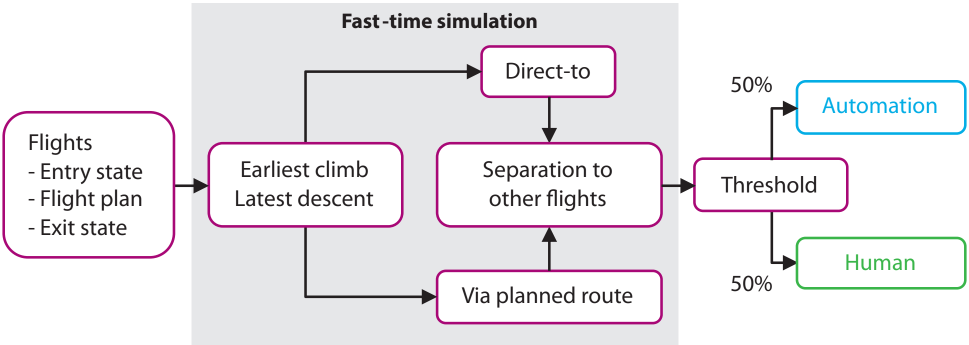


Figure 6.4: Interaction-based allocation scheme with corresponding traffic distributions for the scenario described in Section 6.2.

6.4 Method

6.4.1 Participants

Fourteen professional en-route ATCOs from MUAC, aged between 30–52 years with 6–33 years of professional experience, voluntarily participated. All ATCOs had active licenses for the Brussels sector group. To account for order effects, the participants were split in two groups, balancing for professional experience as much as possible (Table 6.1). All participants provided written informed consent and the experiment was approved by the TU Delft Human Research Ethics Committee under number 3573.

Table 6.1: Participant characteristics.

	Group			
	A	B	Δ	p*
Number of ATCOs	7	7		
Age, years (SD)	42.3 (5.6)	45.0 (6.8)	2.7	0.465
Experience, years (SD)	18.9 (5.2)	22.0 (7.7)	3.1	0.427

* Welch's t-test

6.4.2 Apparatus

SectorX, a TU Delft-built Java-based medium-fidelity simulator was designed to mimic the MUAC interface, to ensure that participants could focus on working with the experimental automation (see Appendix A). A 1920 × 1920 pixels 27″ display was used with a standard computer mouse for control inputs. Eye data were recorded using a Pupil Labs Core head-worn eye tracker and Pupil Capture version 3.5.1 (Kassner et al., 2014). The forward facing out-of-the-world camera recorded at 30 Hz, while the pupils were recorded at 120 Hz. Eight AprilTag markers were placed along the edges of the screen to relate gaze to screen pixels (Figure 6.5). An observer was seated alongside the participant to monitor the eye tracking, make notes and occasionally ask the ATCO to comment on certain decisions or situations.

Aircraft were simulated using the BADA 3.10 performance model (Eurocontrol, 2012). The ATCOs could control heading, route and altitude. Speed and vertical rates were not controllable; flights would follow the default speed profiles from the BADA airline procedures. All clearances were uplinked through datalink, removing the need for voice transmission over radio, and executed by the simulated pilots after a fixed delay of 10 seconds (two radar updates) to simulate the flight crew's response time in processing and implementing a CPDLC instruction. The display could not be adjusted, apart from toggling the history dots and repositioning the dialogs. In terms of support tools, only STCA and MUAC's verification and advisory tool (VERA) were available. STCA would trigger when two flights were expected to have a loss of separation (LOS) within two minutes. VERA could be used to measure the predicted minimum distance between two flights as well as the time to go till that moment, by extrapolating the current speed and direction. It also showed the predicted corresponding locations of both flights. More details about the simulator can be found in Appendix A.

Figure 6.5: Experiment setup, with observer (left) and participant (right) positions.

6.4.3 Scenarios

MUAC provided a traffic sample for Saturday 2 September 2023, consisting of actual radar data and filed flight plans for 6,251 flights (see Section 6.2). All flights not passing through the Brussels sector group were removed. A 25-minute window (17:15–17:40 UTC) was distilled from the data in which the number of flights simultaneously under control by Brussels varied between 25 and 35 (Figure 6.6), totaling to 88 unique flights that would appear on the radar. The base traffic scenario is as shown in Figure 6.2. For training, a 10-minute window was selected that started 15 minutes prior to the measurement scenario (17:00–17:10 UTC), with a slightly lower traffic load of 25–30 flights.

At the time of the selected experiment windows, the sector group was split in four sectors, each staffed by an ATCO duo, consisting of a planner and executive controller. To ensure a sufficiently high workload level, compensating for the lack of voice communication and additional tasks like coordination in the experiment, all Brussels sectors

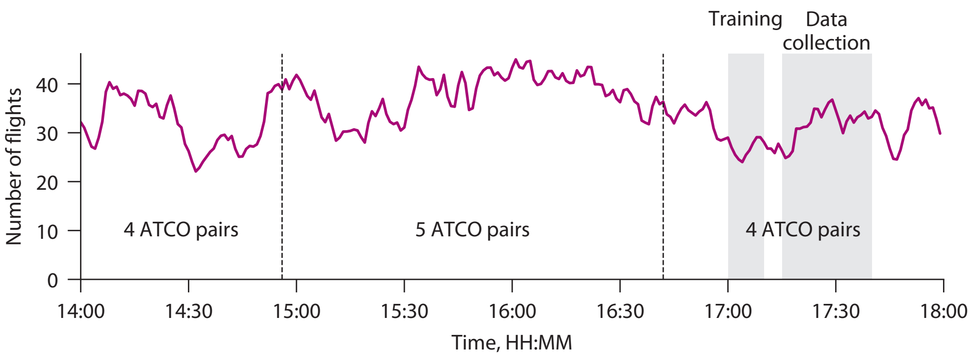


Figure 6.6: Time trace of the number of flights simultaneously controlled by the ATCOs of the Brussels sector group on Saturday 2 September 2023.

were combined into one large airspace. In such a configuration, e.g., during low demand night operations, ATCOs are allowed to handle up to 20 flights for sustained periods and 24 flights during peak moments. As the experiment did not include voice R/T, leading to a workload reduction compared to normal operation, a slightly higher peak of 35 flights was selected. With 50% of the flights automated, the ATCOs would – in theory – never be at their peak capacity. However, due to the expected interactions between human-direct and automation-directed flights, it was expected that the latter flights cannot be completely disregarded from these numbers.

All flights spawned 10 minutes before they were assumed in the data sample and received scripted clearances to prevent pre-sector conflicts as much as possible. Since no R/T was simulated with pilots announcing their entry, the ATCOs were instructed to assume flights at a realistic distance from their sector. One minute after leaving the sector, transferred flights disappeared from the screen to prevent interference with incoming flights, as they were no longer controlled (neither by the ATCO nor the automation). Flights descending out of the airspace where automatically cleared to FL200 after being transferred at FL250 by the ATCO or automation. To prevent additional uncertainty, an international standard atmosphere was used with no wind.

6.4.4 Procedure and participant tasks

All participants followed the procedure outlined in Figure 6.7. During the initial briefing they signed a consent form and were informed about the content and aim of the study, and that they could withdraw from the study at any time. Throughout the experiment, participants were responsible for maintaining separation between flights allocated to them, as well as with respect to flights allocated to automation. Furthermore, they had to ensure that their flights exited the sector at the assigned exit points and flight levels.

In a 10-minute training run, participants were actively encouraged to try all commands and tools available to them. The flight allocation in this run was randomized to not confuse the participants with yet another allocation scheme that was not tested in the experiment. Special attention was given to slight differences in input actions, the lack of letters of agreement with adjacent sectors, and the notion that flights would generally descend at high rates with no ATCO control over vertical climb and descent rates.

After the training, all participants experienced the same measurement scenario twice: once with the flow-based flight allocation and once with the interaction-based

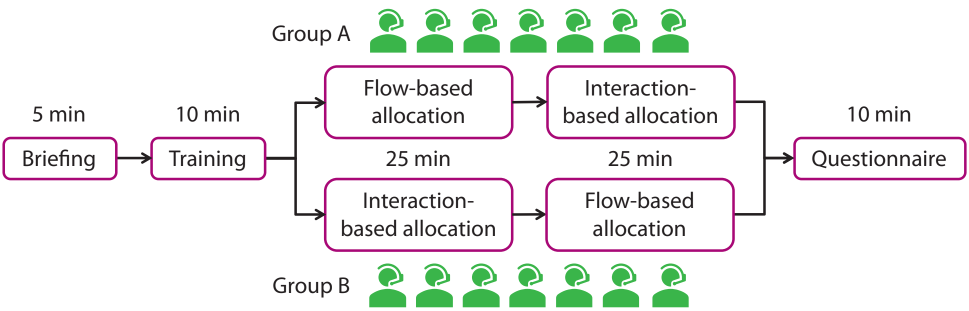


Figure 6.7: Experiment procedure.

allocation. To account for order effects in this within-participants design, participants were split over two groups, each starting with one of the allocations, followed by the other. While the identical traffic samples could evoke scenario recognition (see Appendix C), the different allocations and associated flight colors and control actions were expected to reduce this to an acceptable level.

6.4.5 Automation

During the exercise, the ATCOs were accompanied by an automated 'colleague' that was capable of performing the following tasks on its allocated flights:

- Accept and transfer flights three minutes before entering or leaving the sector.
- Solve conflicts between automated flights by issuing altitude clearances only, ensuring sufficient separation (5 NM, 1,000 ft), plus an additional buffer of 2 NM.
- Deliver flights at their exit point and transfer level, climbing as soon as possible and descending as late as possible.
- Give flights a direct-to towards the route point closest to the exit point for which the path is free of conflicts in the next 8 minutes and does not pass through an adjacent sector.
- Inform ATCOs of mixed conflict pairs 8 minutes before LOS (6 minutes before STCA) through VERA.

The automation used a look-ahead time of 8 minutes, extrapolating each flight's current track, speed and vertical rate, to assess the safety of its clearances before issuing them, taking into account both human-directed and automation-directed flights. This meant that automation would not actively clear a flight into a conflict, but human-automation conflicts could still occur. Either because they were outside the look-ahead window, a flight changed direction while following its route, or the ATCO issued a new clearance to one of their flights. In case of such a mixed conflict, it was up to the ATCO to solve it, under the presumption that automation would not know the ATCO's intents and should not 'fight' for a solution. Automation did apply VERA to inform the ATCO that it had detected a potential conflict that it would not resolve, if the conflicting pair was detected within its look-ahead time. Note that, unlike the ATCOs, the automation could not put flights on a heading. The ATCOs could not take manual control over automated flights, meaning that suboptimal solutions may be needed to solve the conflict.

Automation would only clear flights to a flight level for which it predicted no conflicts within its look-ahead time. This prevented the automation from blocking an excessive altitude band for other flights. The automation would issue the next clearance when a flight was within 3,000 ft of the previously cleared level, to prevent continuous clearances, as the look-ahead window moved ahead.

Apart from showing the uplinked clearances in the flight labels, automation did not provide any feedback on its intent. The simple rule-based approach was designed to support trust buildup (Lee and See, 2004) and reduce the need for more extensive communication, although previous experiments have shown that ATCOs like to be able to see information such as the expected top of descent (Chapter 3). Adding and visualizing this was, however, outside the scope of the experiment.

6.4.6 Independent variable

There was one independent variable: the two distinct allocation schemes discussed in Section 6.3, with traffic densities as visualized in Figure 6.8. To account for ordering effects, the participants were divided over two groups with distinct independent variable orders. The color of the flight label and radar symbol indicated which flights were allocated to the human or the automation. Unlike the experiment from Chapter 3, the human-automation allocation was fixed per scenario and could not be adjusted by the ATCO.

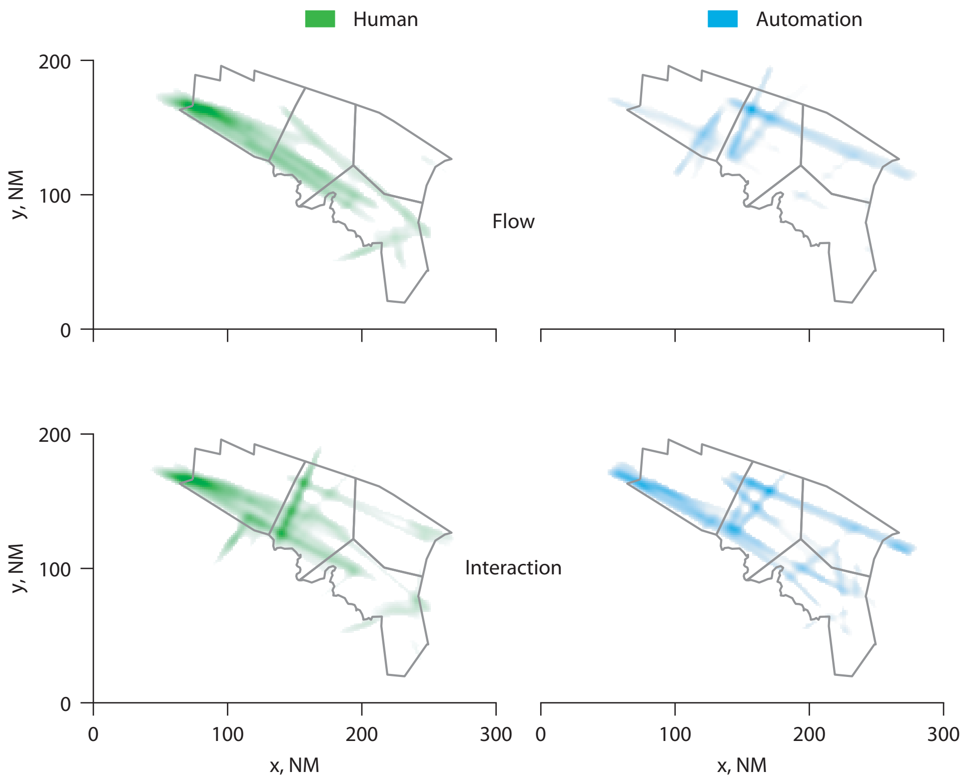


Figure 6.8: Traffic density maps of the two scenarios, split per agent.

6.4.7 Control variables

The following control variables were constant for each participant and both scenarios:

- Traffic sample, as described in Section 6.4.3,
- Atmospheric conditions: international standard atmosphere without wind,
- Automation capabilities, as described in Section 6.4.5,
- Pilot delay: 10 seconds,
- ATCO support systems: only VERA and STCA.

6.4.8 Dependent measures

The following measures were collected in or derived from the simulation:

- *Perceived workload:* Measured through an instantaneous self-assessed (ISA) rating on a 0–100 scale every 5 minutes (Tattersall and Foord, 1996).
- *Support tool usage:*
  - Flight pairs for which VERA was activated more than 100 ms, either by the automation or the ATCO.
  - Lengthening of speed vectors beyond 1 minute.
  - Triggering of STCA.
- *Gaze patterns:* On-screen gaze locations provided by the eye tracker.
- *Control activity:* The number, type and timing of issued clearances (altitude, heading and direct-to) and instructions (assume and transfer).
- *Efficiency:* Track miles of flights, as described in Section 6.4.9.
- *Post-experiment questionnaire:* With open and Likert-scale questions on the functioning of the automation, two allocation schemes and simulator fidelity.

6.4.9 Data analysis

Automation was not able to clear flights beyond their XCOP, but ATCOs are used to do so in their daily operation. Many of these points are on a more or less straight path beyond the XCOP and thus have only a limited effect on the traveled distance. To nevertheless ensure a fair comparison, the XCOP was projected as XCOP* on a line between the cleared waypoint and the current aircraft position, at the same distance from the cleared point as the XCOP (Figure 6.9). Secondly, variations in transfer locations meant that the remaining distance towards the XCOP (or XCOP*) had to be added to the traveled distance since the start of the scenario for flights that had not yet reached that point by the end of the scenario. For any other flight the track miles were calculated until reaching the minimum distance to the XCOP.

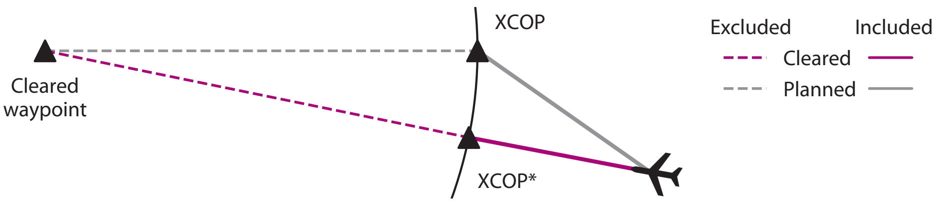


Figure 6.9: Calculation of track miles for flights that are cleared to points beyond the XCOP.

6.4.10 Hypotheses

It was hypothesized that the use of the interaction-based allocation scheme:

H1 would lower the perceived workload due to the reduced interference with automation-directed flights.

H2 would reduce VERA usage on mixed flight pairs, as fewer mixed conflicts should occur.

H3 would reduce gaze activity on flights allocated to the automation, as these flights become less relevant for the flights allocated to the human.

H4 allowed for more efficient conflict resolution, in terms of track miles, by assigning both flights in a (predicted) conflict to the same agent, giving that agent more freedom over how to solve the conflict.

H5 was overall more appreciated by the ATCOs as it would give them more freedom to implement their own plan and it would minimize the number of mixed conflicts for which the automation's capabilities and actions had to be considered.

6.5 Results

All eye tracking data with confidence levels below 0.9, as reported by Pupil Player version 3.5.1, were excluded. Results are presented in accordance with the dependent measures defined in Section 6.4.8 and preceded by a general description of the flight distribution between human and automation and subsequent exclusion of part of the data.

6.5.1 Flight distribution and data exclusion

Figure 6.10 shows the distribution between manual and automated assumed flights, averaged over the fourteen ATCOs. The freedom to assume or transfer a manual flight was entirely with the ATCOs, which meant that the number of flights under control at a specific timestamp varied up to five flights between any two ATCOs. The flight share was comparable between the two allocations for the first 20 minutes. With 48% and 45% manual flights for the flow- and interaction-based allocation, respectively, it was close to the 50% division that was targeted in the experiment.

In the final five minutes of the flow-based scenario, the share of manual flights increased to 63% on average, versus 46% for the interaction-based allocation. This difference was mostly caused by a bunch of flights from London that all had to climb more than 5,000 ft and were thus allocated to the ATCO in the flow-based scenario. As the number of assumed flights is directly related to measures such as the number of issued clearances, checked conflict pairs and potentially workload, the remainder of this analysis focuses on the first 20 minutes of each simulation run, unless otherwise noted.

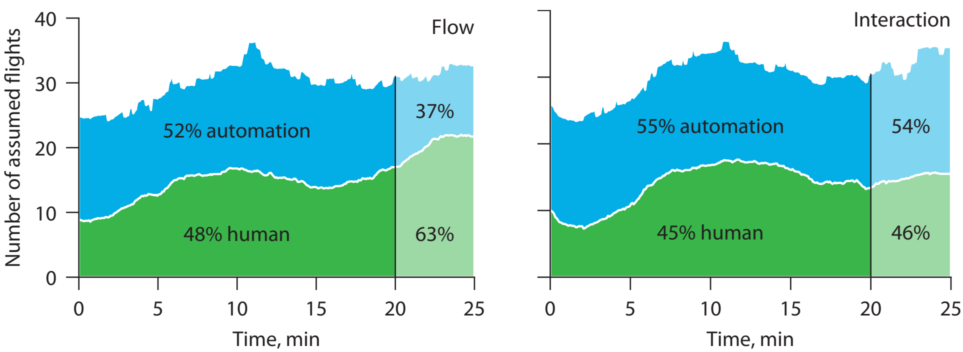


Figure 6.10: Stacked time traces of the number of assumed flights per agent, averaged over all ATCOs. The final five minutes were excluded from the rest of the analysis.

The distinct flight allocations were clearly reflected in the location at which flights were assumed by either the ATCO or the automation (Figure 6.11). With the interaction-based scheme, in- and outbounds to and from Amsterdam are manually assumed, while they were automated with the flow-based allocation. Whereas the automated flights were always assumed at the same location, the manual flights show a considerable spread between participants due to the absence of pseudo-pilots calling in.

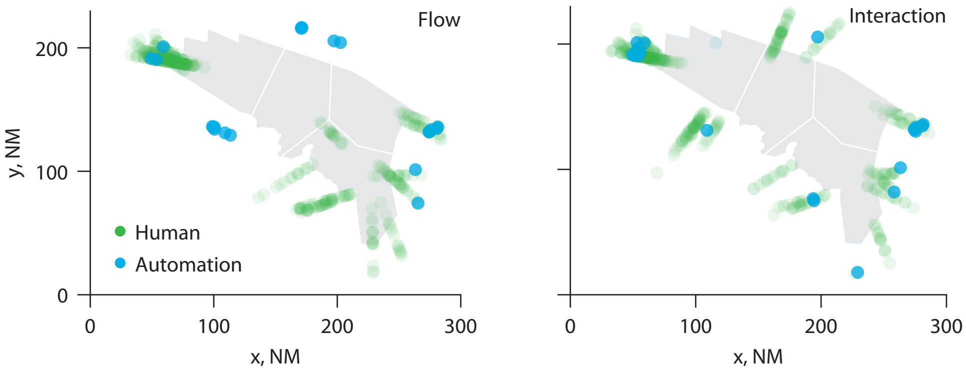


Figure 6.11: Location at which flights were assumed. Each flight appears fourteen times, once per ATCO.

6.5.2 **Perceived workload**

In the self-reported workload rating (Figure 6.12), a downward trend is visible after the first 10 minutes for the interaction-based allocation that is not replicated with the flow-based allocation and only marginally reflected in the number of manually assumed flights. The lower number of assumed manual flights in the final five minutes of the interaction-based scenario is, however, clearly reflected. Only these final five minutes led to a statistically significant higher workload in the flow-based allocation (M $= 0.231$, SD $= 1.174$) than in the interaction-based allocation (M $= -0.717$, SD $= 0.972$): $t(13) = 3.463, p = .004$, with a large effect size (Cohen's $d = 0.93$). The relatively large difference can be entirely attributed to participant Group A, who first experienced the flow-based scenario, as Group B showed a more symmetric workload distribution.

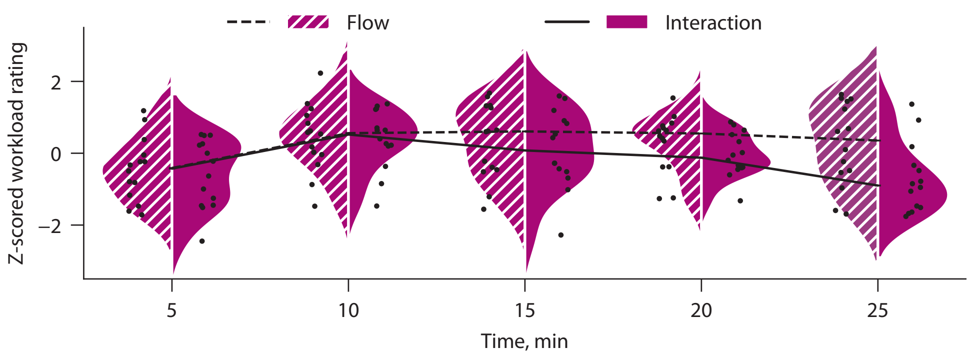


Figure 6.12: ISA workload ratings, Z-scored per ATCO. Lines connect median values per allocation.

6.5.3 **Support tool usage**

VERA

As one of the primary tasks of an ATCO, timely conflict detection is a key performance indicator. On average, 19 unique flight pairs were checked with VERA to verify whether the selected pair was in conflict. One ATCO was a noticeable outlier and checked only one pair. Figure 6.13 shows the VERA actions initiated by the ATCOs as well as the automation. Note that the latter could only add mixed conflicts and would automatically remove mixed pairs from VERA when the flights did not have a predicted LOS, even though the ATCOs may still have been interested in monitoring the pair. To prevent counting repetitive activations, only the first VERA action for each unique flight pair is included in the figure. Furthermore, only pairs for which VERA was active for more than 100 ms are considered, as VERA was, once attached to a flight, (temporarily) added to any other flight that the mouse passed over while it was moved towards a flight of interest by the ATCO.

The total number of VERA actions did not differ per allocation scheme (18 on average), but there was a shift from mostly mixed flight pairs in the flow-based scenario to a more balanced share between mixed and manual pairs in the interaction-based scenario. According to a Wilcoxon signed-rank test with Bonferroni correction, both the increase in manual pairs ($Z = -3.180, p = .002, r_B = -1$) and the decrease in mixed pairs ($Z = 3.170, p = .002, r_B = 0.962$) were significant. The rank-biserial correlations suggest that these are both large effect sizes. The significant shift is a testimony to the interaction-based allocation's functioning in reducing the number of complex mixed conflicts, by assigning interacting flights to a single agent. Furthermore, in the flow-based scenario, slightly more automated pairs were second-guessed, signaling a lack of

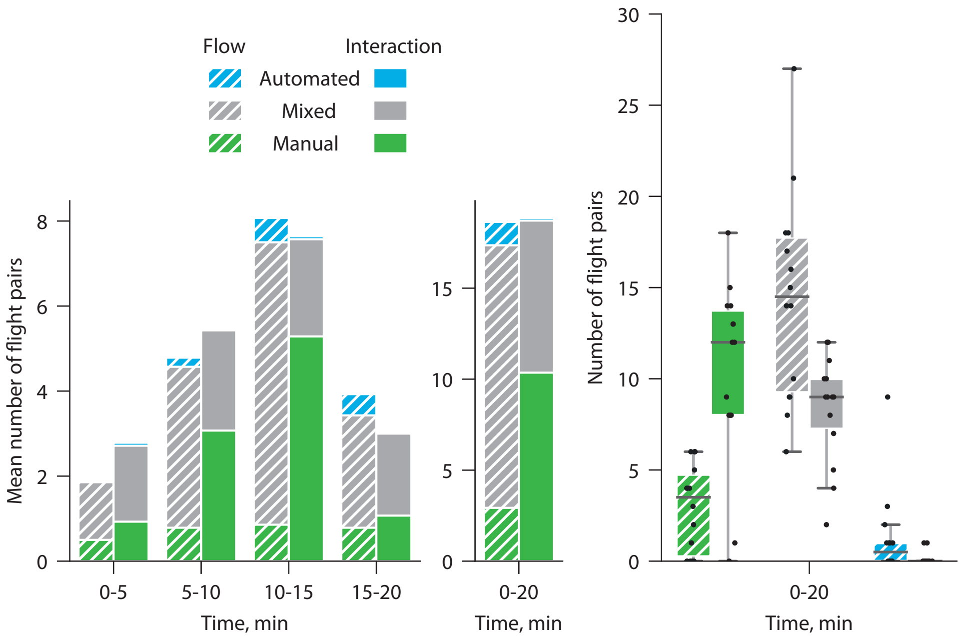


Figure 6.13: Number of unique flight pairs per ATCO for which VERA was activated.

trust in the automation. This was especially true for one particular ATCO who checked nine mixed pairs in the flow-based allocation. It must be noted that in some cases these pairs were merely checked to confirm and/or highlight the severity of – according to the ATCOs – a bad decision by the automation.

Speed vector

Next to VERA, the ATCOs could increase the length of the speed vectors beyond the default one minute to extrapolate the positions of flights. This was sparsely used with only 38 activations in the entire experiment and was not significantly different between the two scenarios. ATCOs mostly used this to judge whether a given lateral clearance provided sufficient future separation. VERA on the other hand provides a more accurate measurement of the CPA and a way to keep track of potential conflicts over a prolonged time. Since the speed vector length was changed for all flights at once, it is not possible to reliably link this to specific flight pairs. It is worth noting though, that the sparse use of VERA by some ATCOs is only partly compensated by a higher use of speed vectors.

STCA

When ATCOs fail to detect and/or respond too late to conflicts, an automatic STCA can trigger. This occurred eight times during the experiment, all in the flow-based scenario and each with a different ATCO. The primary cause of STCA was the late descent by automation of flights inbound to Amsterdam and subsequent conflict with a human-controlled west-bound flight that would never have happened in current operations. The six ATCOs that did not get the STCA proactively steered their (manual) flights around this particular conflict area to ensure sufficient separation.

6.5.4 Gaze patterns

The eye tracker allowed to record the ATCOs' gaze patterns. Figure 6.14 shows the areas with highest gaze density, as well as the areas with highest traffic densities, for all ATCOs combined. For clarity, the traffic density figures are split per controlling agent, while the gaze density contours are duplicated. Traffic densities are based on flight trajectories until the flight was transferred to the next sector, as transferred flights are generally not of interest to ATCOs and rarely looked at. Furthermore, about 16% of the gazes were collected while a clearance menu was open. These have been filtered out to prevent incorrectly mapping them to flights on the screen that may have been obfuscated by the menu.

The large concentration of flights entering the sector from the west is clearly visible. In both scenarios, an area of high gaze density follows the east-bound transition of a cluster of green flights. Two additional observations stand out. First, areas with high gaze density generally coincide with areas with high densities of manual traffic. Areas with high density of automated flights only seem to coincide with high gaze activity when they overlap with dense manual areas. Second, the gaze patterns are largely comparable between the two scenarios. This is not surprising, as the sector and traffic streams were identical, meaning that hotspots occurred at similar locations.

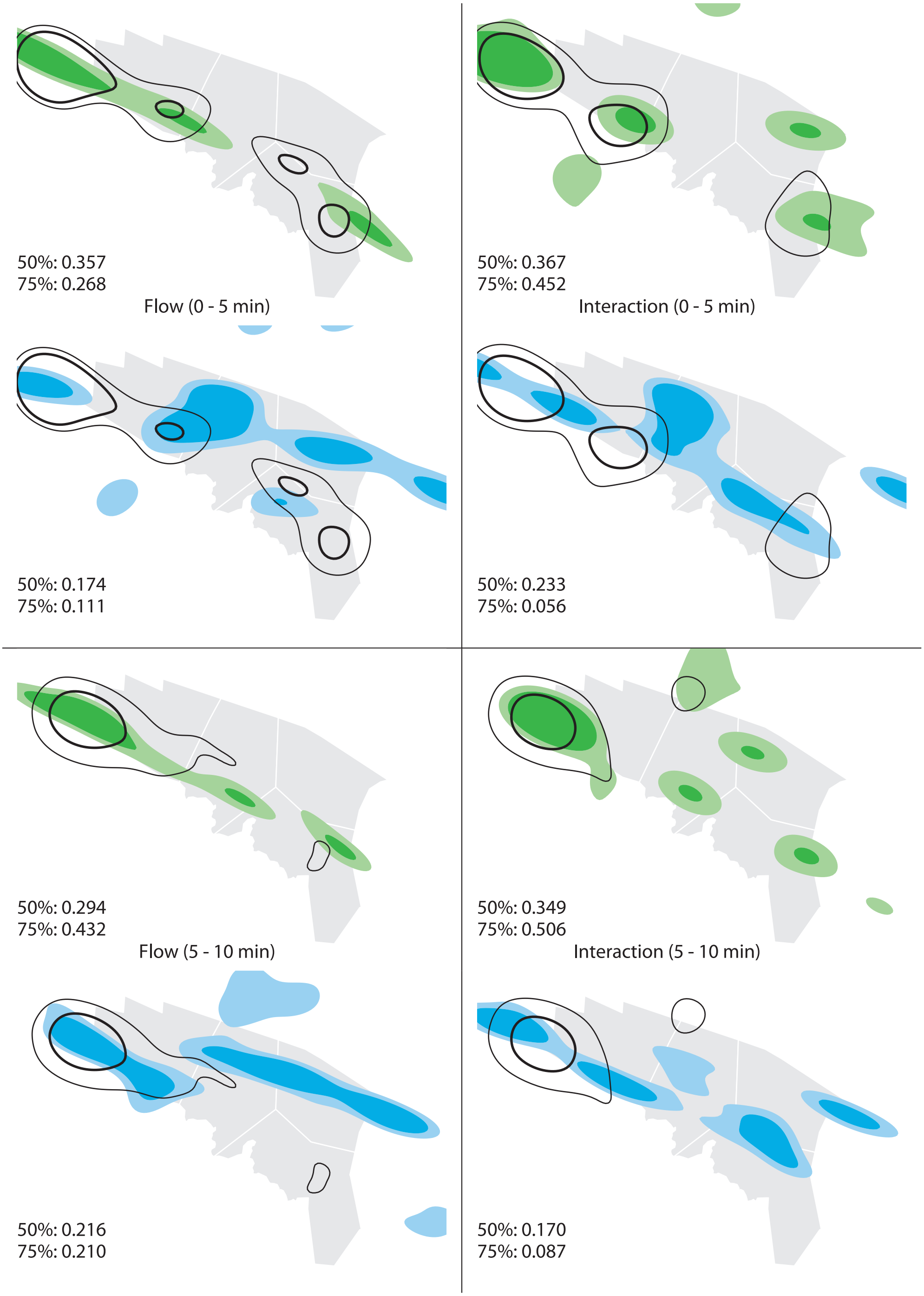
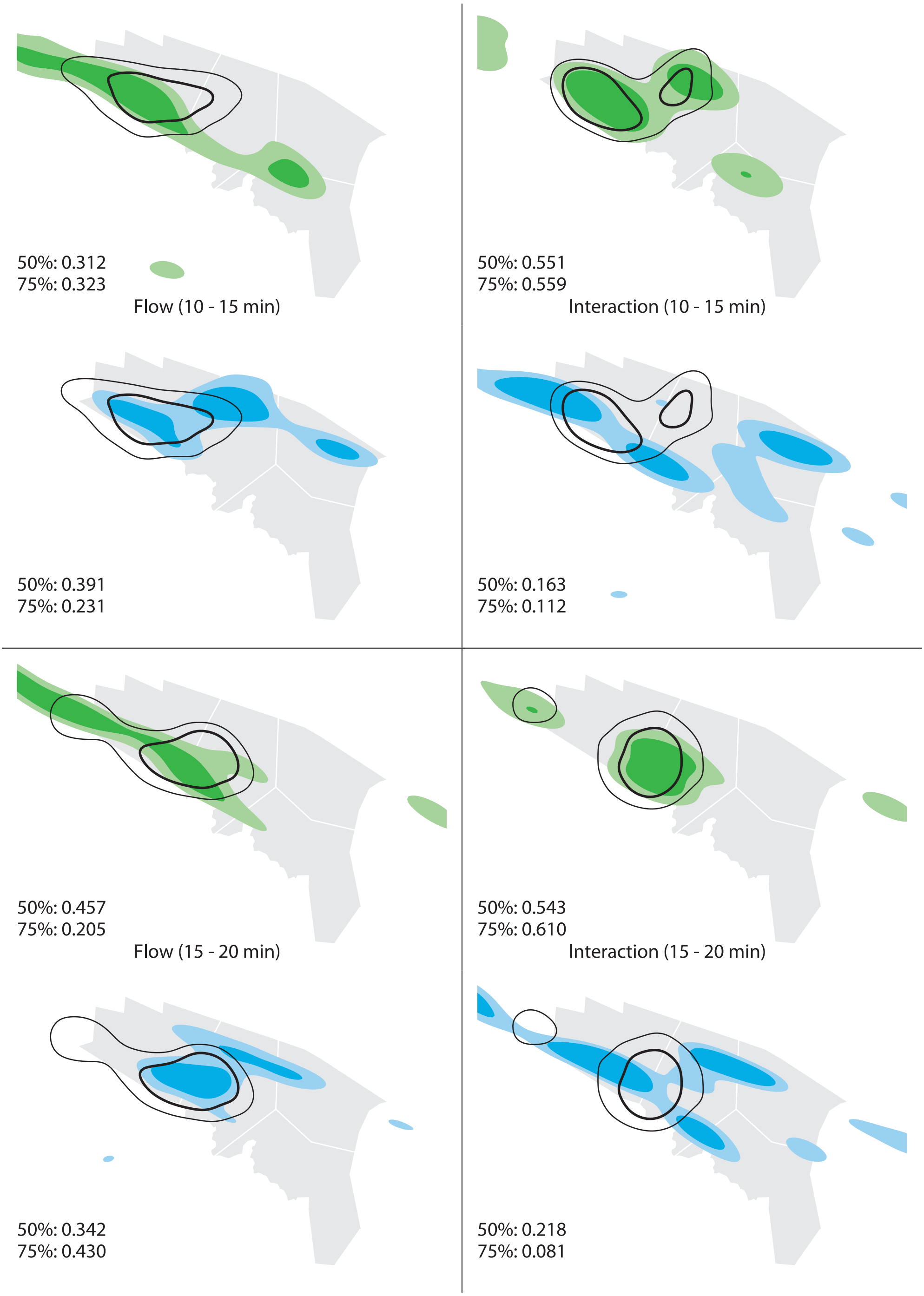


Figure 6.14: High density traffic (filled polygons) and gaze (outlines) areas, for all ATCOs combined, over a 50% and 75% kernel density. To quantify the match between gaze and traffic, Jaccard indices are given. Green refers to human-controlled flights, while blue relates to automation-controlled flights.



6

Figure 6.14: Continued from previous page.

To quantify these observations, the Jaccard index is calculated per Equation (6.1), i.e., the 'similarity' between high density areas of radar blips and ATCO gazes.

$$\text{Jaccard index} = \frac{\text{Area of overlap}}{\text{Area of union}} \tag{6.1}$$

Shown in Figure 6.15, the interaction-based allocation resulted in higher Jaccard indices for green flights than the flow-based allocation throughout the scenario, while the blue flights have a lower Jaccard index, except for the 50%-index between 0–5 minutes. With the interaction-based allocation, there is a clear 'gap' between the Jaccard indices for human and automation-directed flights, an indication for the parallelism of the system. The gap is much reduced or even absent in the flow-based allocation, with the automated flights sometimes even having a better match with gaze than the human-directed flights. To illustrate, at 15–20 minutes in Figure 6.14, the primary gaze is clearly on human-controlled flights in the interaction-based scenario, while it is more aimed at blue flights in the flow-based scenario. This is reflected by a 'reversal' of the 75% Jaccard indices in Figure 6.15 (blue line above the green line).

The lower half of Figure 6.15 shows per-ATCO results. While the indices for the human-controlled flights are somewhat lower, the clear gap with automation-controlled flights is still present in the interaction-based scenario. The lower spread in this scenario's results indicates that ATCOs had more uniform gaze patterns.

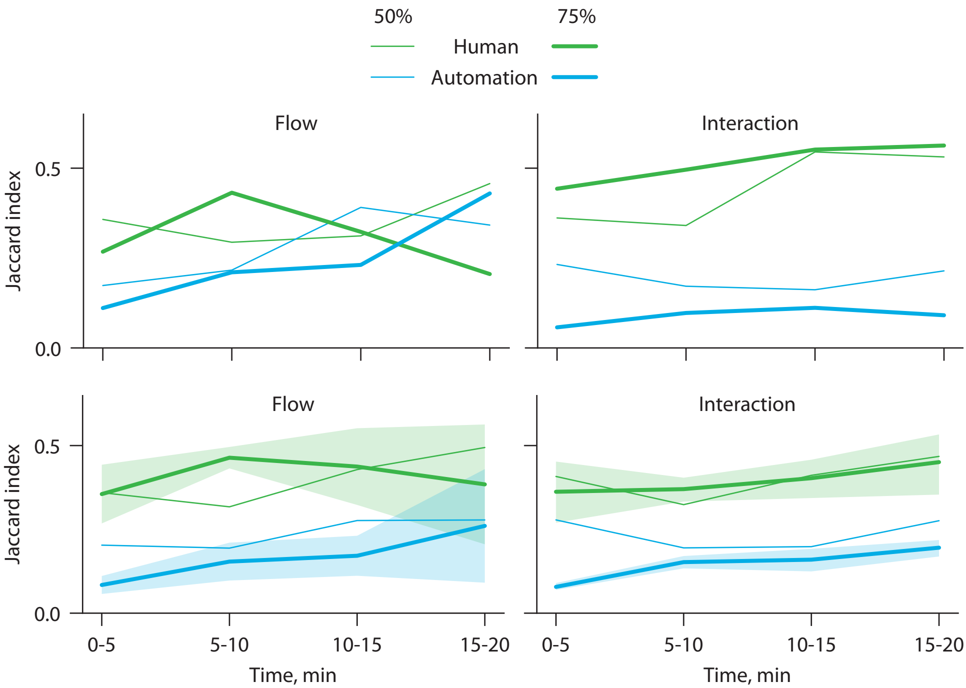


Figure 6.15: Jaccard indices over time. Lines in the top half of the figure correspond to Figure 6.14 with all ATCOs combined, while the lower half shows the average when calculating Jaccard indices per ATCO. The 75%-indices are shown there with 95% confidence intervals to illustrate between-ATCO variance.

6.5.5 **Control activity**

To resolve conflicts and/or make flights meet their exit conditions, the ATCOs issued on average a total of 61 and 67 manual clearances for the flow- and interaction-based scenario, respectively, while automation issued 53 and 48 clearances. Since different flights require different actions, the aggregated distribution between automation-issued and ATCO-issued clearances cannot be readily compared, but the differences per agent between scenarios can shed some insights.

Eleven of the ATCOs issued more altitude clearances in the interaction-based scenario than in the flow-based scenario (Figure 6.16), which was a significant change according to a Wilcoxon signed-rank test ($Z = -3.18, p < .01$). Altitude clearances also show the largest difference of all clearances between the two scenarios, with the ATCOs on average issuing 41 in the interaction-based scenario and only 33 (-22%) in the flow-based scenario. This was only partially compensated by automation issuing, on average, four altitude clearances more (28 versus 32). With the same entry and exit conditions, this means that more intermediate levels were given in the interaction-based scenario by the ATCOs.

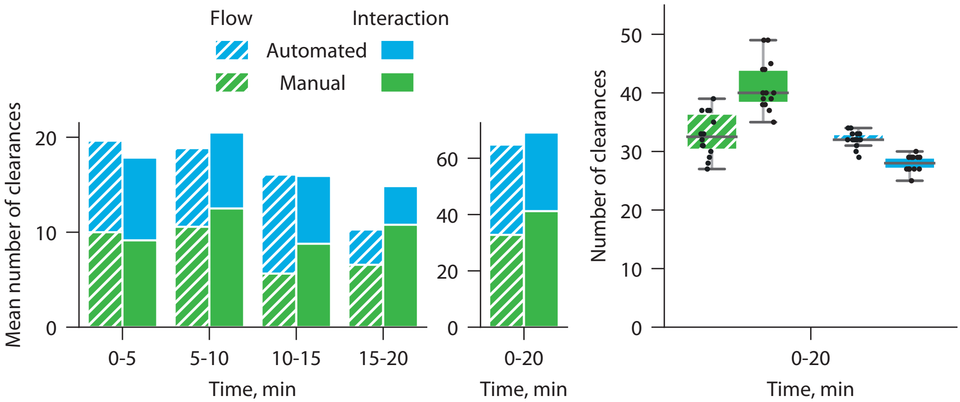


Figure 6.16: Number of altitude clearances per ATCO.

Figure 6.17 shows that the number of manual direct-to's was similar between the two scenarios, and that on average the flow-based scenario elicited two additional direct-to's: one by either controlling agent. Between 10 and 15 minutes into the scenario, the number of automated direct-to's was substantially lower in the interaction-based scenario (by five clearances on average), which was not offset by an increase in manual direct-to's or heading clearances. In addition, many ATCOs put one of their flights on a heading during this period to steer around an automated inbound to Amsterdam in the flow-based scenario (Figure 6.18). The number of issued heading clearances over the entire scenario shows substantial variation between ATCOs, ranging between 1 and 24. These numbers are not directly proportional to the number of direct-to's, even though most flights on a heading received a direct-to clearance at some point to return to the planned route. Only the difference in automated direct-to's was statistically significant, but at only one flight it can be considered negligible.

6

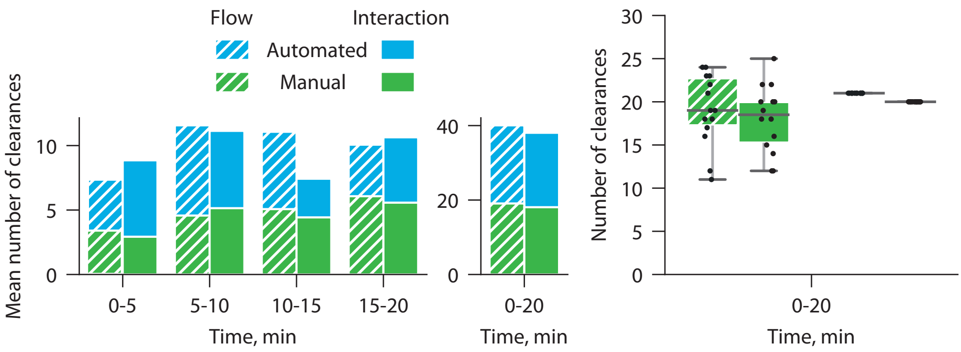


Figure 6.17: Number of direct-to clearances per ATCO.

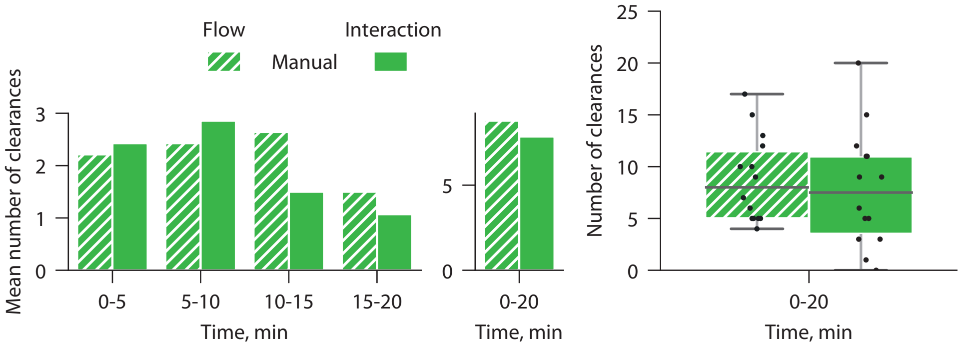


Figure 6.18: Number of heading clearances per ATCO.

The location where clearances were issued was slightly different in the two scenarios (Figure 6.19). Especially the in- and outbounds to/from Amsterdam are clearly visible in the interaction-based scenario as green stripes, corresponding to increased gaze activity in that area (Figure 6.14). Also note that the automation was very consistent, whereas the ATCOs timed their clearances on an individual basis leading to a much larger spread.

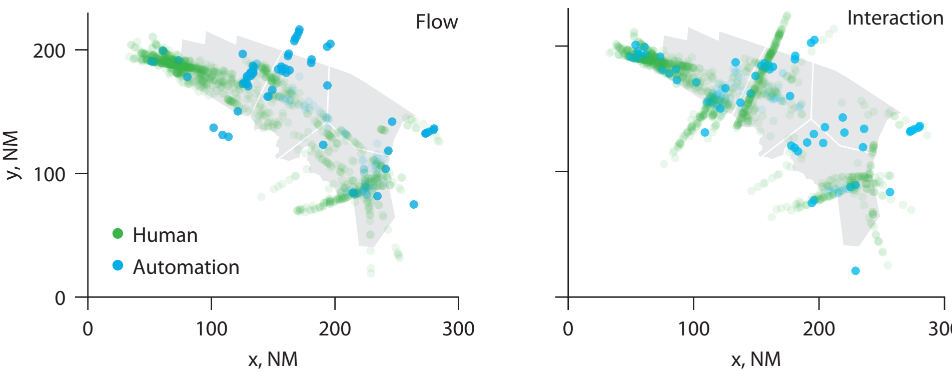


Figure 6.19: Location at which clearances (excluding assume and transfer) were issued to flights. All ATCOs are included, so a similar clearance to a single flight can be shown in multiple locations.

6.5.6 **Efficiency**

The total system efficiency can be related to the total track miles traveled by all flights, with a lower track distance resembling fewer en-route delays, lower fuel consumption and higher overall efficiency. As the traffic sample and environmental conditions were identical in both scenarios, the two allocation schemes can be compared. In 186 (27%) out of 693 direct-to clearances, the ATCO cleared a flight to a route point beyond the XCOP. This is standard practice, but the ATCOs were briefed not to do so in the experiment, unless required for the safety of the traffic. The procedure discussed in Section 6.4.9 compensates the track mile calculation for the variety in last cleared points.

In all, the total flown distance was slightly lower in the interaction-based scenario for nine ATCOs (M = 4.0 NM, SD = 3.3 NM) and higher for five ATCOs (M = 3.0 NM, SD = 2.1 NM). These differences are not significant on a combined distance of circa 13,400 NM for all flights in a scenario. However, since many small shortcuts can compensate for one big delay, the track mile difference per flight might be more relevant from an airline perspective. Including every of the 80 unique flights from the first 20 minutes of the traffic sample once per participant yields 1,120 flights (80 x 14) for which a track difference can be calculated.

For display clarity, Figure 6.20 only shows the 178 flights with an absolute track difference of at least 0.5 NM. All 266 flights exclusively allocated to the automation in both scenarios flew consistent track miles and were among the 942 excluded flights. Flights on either side of the ‘excluded’ bar had shorter tracks when one of the two allocation schemes was applied, for which the color indicates which agent was responsible for the flight. The figure shows that out of the 294 flights that were allocated to the ATCOs in both scenarios, 36 flew fewer track miles in the interaction-based scenario and 15 in the flow-based scenario (the rest was excluded). This suggests that the ATCOs were able to implement more efficient routing for some of their manual flights, now that they were not hindered by automation-controlled flights.

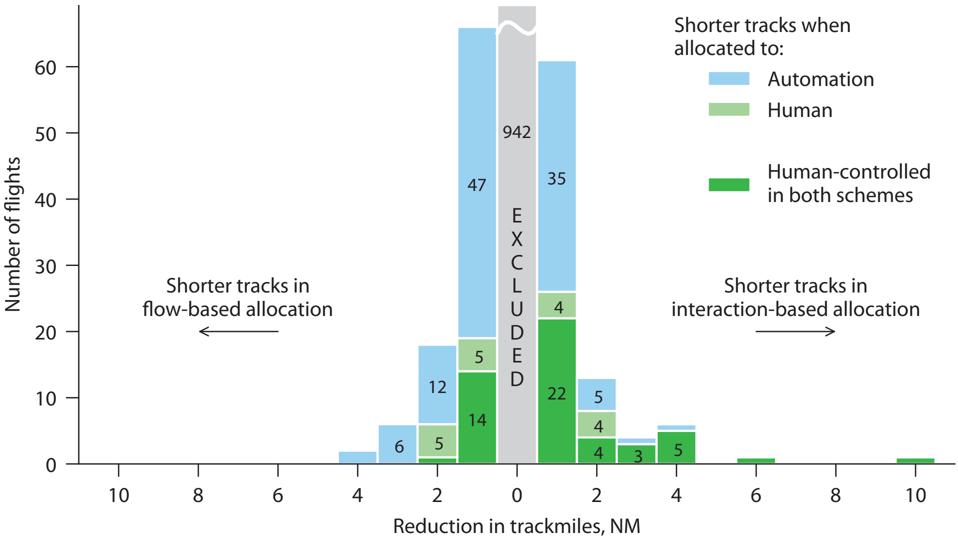


Figure 6.20: Differences in track miles per flight between the two allocation schemes.

On the contrary, 109 flights that were allocated to automation in only one of the two scenarios flew longer distances when allocated to the ATCOs in the other scenario and only 18 flights flew shorter distances under human control. Figure 6.21, shows three examples of the former with their respective planned routes and XCOPs. The flight towards GTQ flew close to the sector boundary, meaning that it could not yet turn towards its exit point without clipping the adjacent sector. The automation in general issued the turning clearance earlier than its human counterparts, resulting in a slightly (ca. 2 NM) shorter track distance. One ATCO cleared the flight to a point beyond the XCOP, resulting in a more south-bound trajectory, crossing that of the automated flights. The flight towards PITES saw considerable variation in when ATCOs cleared it towards PITES, with all but one ATCO turning later than the automation. Finally, the flight towards COA crossed the inbound stream to Amsterdam, which the automation in the flow-based scenario descended later than the ATCOs did in the interaction-based scenario, resulting in many ATCOs steering their flight around these flights. Four ATCOs took timely evasive action, while the rest waited until the STCA triggered as indicated by the dense cluster of turning points. Also note that only three ATCOs chose to turn left to pass behind the conflicting flight. The wide variety in end-locations of these flights exemplifies the need for the track mile calculation as shown in Figure 6.9.

6

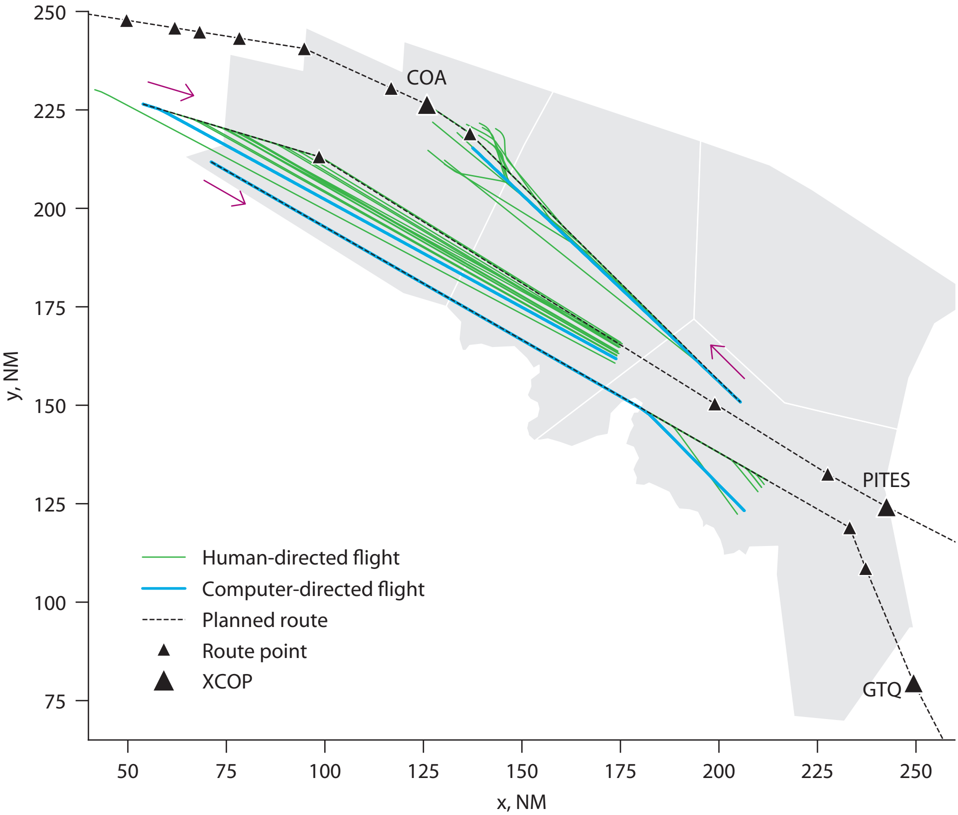


Figure 6.21: Ground trajectories of three flights for both scenarios and all ATCOs. Trajectories are clipped from being assumed until transfer to the next sector or end of the scenario.

6.5.7 **Post-experiment questionnaire**

Flight allocation

In general, more ATCOs liked the interaction-based allocation (Figure 6.22), with none of them disliking that allocation a great deal, versus five ATCOs disliking the flow-based allocation a great deal. Two ATCOs preferred the flow-based over the interaction-based allocation, and three were indifferent as to which allocation was used.

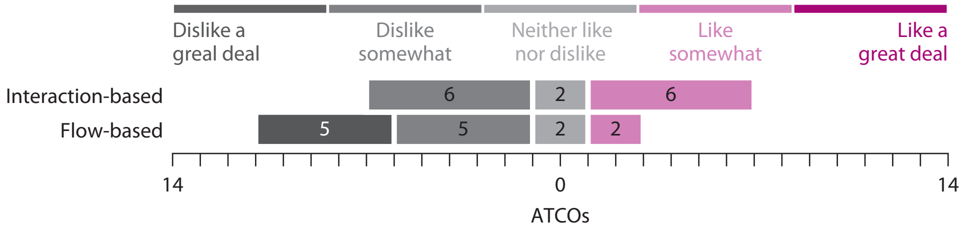


Figure 6.22: ATCO opinion about experimental allocation schemes.

The interaction-based allocation scored generally favorable on all aspects of the post-experiment questionnaire (Figure 6.23). All ATCOs who reported a lower workload with the flow-based allocation were in Group B, meaning that they had first experienced the interaction-based allocation. Presumably their increased experience with the automated 'colleague' and the traffic scenario played a role here. The other aspects show no clear differences between the two groups. On safety and efficiency, the majority of ATCOs were indifferent as to which allocation was better, with the rest having a slight preference for the interaction-based scheme. Several ATCOs mentioned that they see safety as binary, it is either safe or not safe. Since automation would issue, according to current ATC rules, a number of unsafe clearances in both allocations, the ATCOs considered both scenarios 'equally' unsafe.

In an open question, the ATCOs were encouraged to think about what they would consider when designing an allocation scheme. Apart from one ATCO who outright rejected *"working live traffic alongside artificial intelligence"*, all ATCOs mentioned that conflict-free flights that do not require any (planned) action can indeed be automated. Six ATCOs proposed a maximum flight level change for automated flights, ranging from 0 to 2,000 ft (matching one of the proposed allocation schemes in Chapter 3), or simply *"the less vertical movement, the better."* In the horizontal plane, turns can be treated similarly, as they complicate extrapolation of the predicted flight path. Assigning flights that interact with each other to the same agent was mentioned by three ATCOs, on the premise that this allows the selection of the most optimal solution, regardless of whether that is from an airline, environmental or ATCO perspective. One ATCO suggested that flights with subjacent exits (i.e., descending to FL250 within the airspace) should not be delegated to automation at all, as these often have stricter exit conditions and frequently require coordination with the ATCO responsible for the lower sector. Bunches of traffic should be given to the ATCO, while clearly defined trails of traffic (e.g., EHAM outbounds) can be given to the automation, provided that their climb profile is monitored.

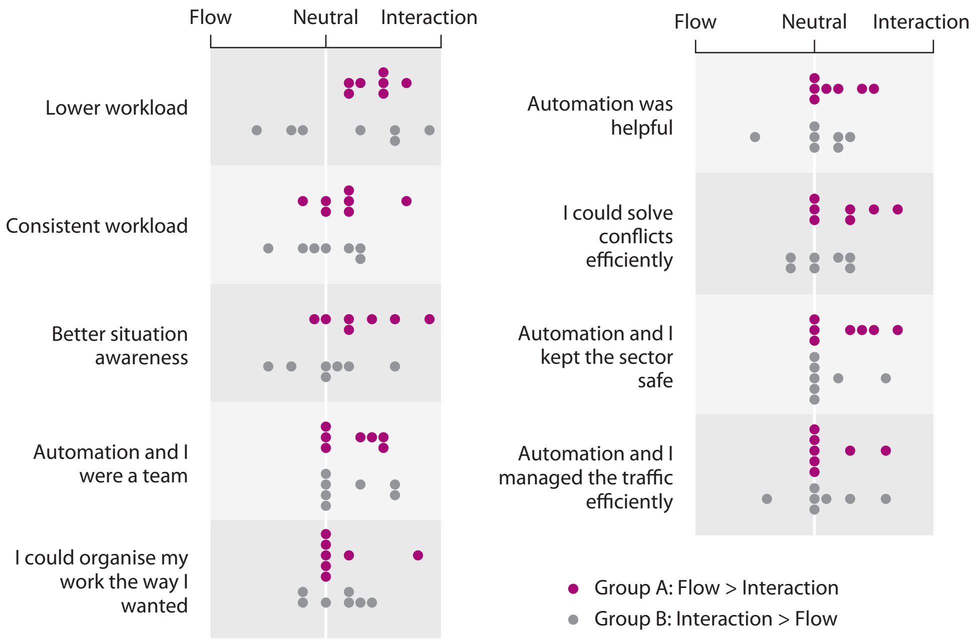


Figure 6.23: Relative subjective differences between the two allocation schemes. Dots towards either extreme indicate a strong(er) preference for one scheme over the other.

6

Automation

Although the design of the automated agent was not explicitly investigated, some results do give an insight into the ATCO's stance. Two ATCOs mentioned, after the experiment, that they lacked sufficient knowledge about the automation's capabilities due to the limited training they received. An interesting example was when an in-trail automated flight was overtaking a leading flight that was also automated. Both flights were flying level at their (identical) TFL. Automation solved this conflict by lowering the trailing flight one flight level, which surprised some ATCOs who had not (yet) identified the conflict.

As shown in Figure 6.24, ATCOs were divided on whether the automation was reliable. Most ATCOs thought the automation was not predictable, but at the same time quite consistent, as was already proven objectively by the small variances in number of automated clearances (Figures 6.16 and 6.18). The ATCOs were also ambiguous about how much trust they had in the automation. One of them phrased it as having *"a lot of trust in the automation between the blue flights"*, but less for mixed pairs where it felt like *"my green flights were often ignored by the computer"*. ATCOs would have liked to be able to 'nudge' the automation to start an action, such as a descent. Especially when a mixed conflict could be most easily solved by the automation starting a planned descent earlier, the ATCOs did not like having to solve the conflict by adjusting their green flight(s). Similar comments were collected in flight-centric trials by Martins et al. (2019), where participating ATCOs in addition mentioned that the corresponding item in the flight label should highlight when a clearance is given by the other ATCO.

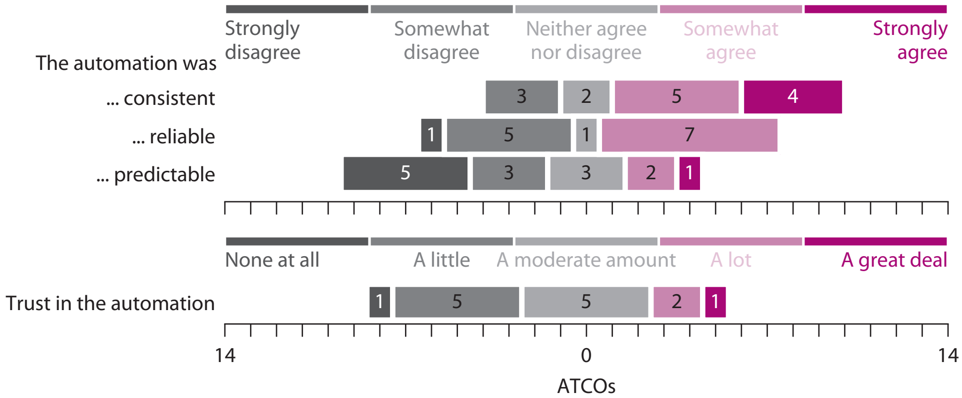


Figure 6.24: Questionnaire results on statements about the experimental automation.

In general, the ATCOs appreciated that automation flagged potential mixed conflicts with VERA. However, some requested the VERA to be mutable, if the ATCO had determined the alert to be non-urgent (e.g., because one of the flights was expected to descend well ahead of the crossing point). The fact that automation could remove a manually-added VERA pair turned out to be an unintended nuisance in the experiment.

Simulation fidelity

Figure 6.25 shows the ATCOs' opinions on the realism of the simulation. Traffic scenarios were considered realistic, although some flights had to exit the sector to an adjacent MUAC sector at unusual levels. This was traced back to the absence of intra-sector TFLs in the flight plan sample that was used to construct the scenario.

The appearance of the interface was considered to largely resemble the real HMI, except for several inputs requiring slightly different actions. For example, right instead of left clicking a label to transfer a flight. Most ATCOs managed to adapt to this behavior during the training scenario, although for many, occasional miss clicks occurred throughout the experiment. This may have added some additional cognitive load.

Aircraft behavior leaned towards somewhat unrealistic, which was mainly attributed to high climb and descent rates, which do occur in reality but are, unlike in the experiment, not attained by default. Flights increasingly climb at vastly different cost indices, leading to greater variance in vertical rates. Being able to issue vertical rates would have

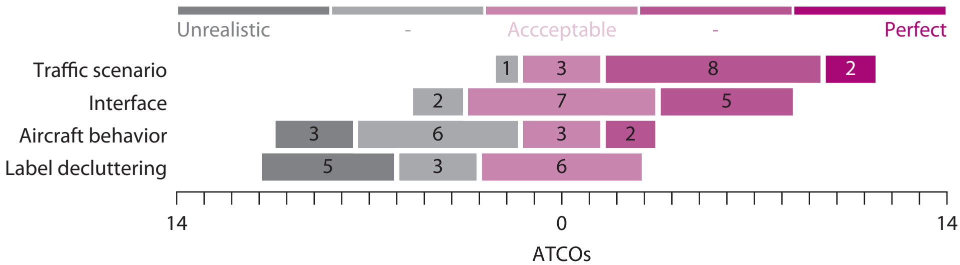


Figure 6.25: Questionnaire results on simulation fidelity.

given more precise control over timely climbs or descend. Furthermore, speed control would have eased resolving in-trail conflicts. As ATCOs are used to regular delays of up to one minute between their clearance and the flight actually implementing it, they considered the fixed 10 seconds pilot delay too short. In some cases this resulted in ATCOs issuing a clearance earlier than they would have done in hindsight.

Finally, label decluttering was considered acceptable when it did a good job, but the occasional presence of overlapping labels was considered unacceptable and is something that the real system reliably avoids. Most overlaps occurred in the western part of the sector with many tightly packed flights entering from the United Kingdom. ATCOs could manually rotate these labels to declutter them.

6.6 Discussion and recommendations

6.6.1 Hypotheses

Overall, the results show a small improvement on all hypothesized aspects with the interaction-based allocation. Both the ISA ratings and the post-experiment questionnaire point to a workload reduction. It can, however, not be ruled out that this was caused by the asymmetric flight allocation in the last five minutes of the flow-based scenario. Hypothesis H1 about the expected workload reduction can therefore only be conditionally accepted. Interestingly, contradicting the lower perceived workload, the number of manual altitude clearances was 26% higher in the interaction-based scenario. ATCOs are probably so used to providing those clearances that it barely adds to their workload, whereas dealing with the novel automation takes more effort and thus becomes the dominant factor. The absence of voice communication may have played an important role here as radio transmissions are known to be a significant contributor to (perceived) ATCO workload (Dow and Histon, 2015).

The shift in VERA usage from primarily mixed conflict pairs to more manual pairs was significant, confirming Hypothesis H2. This result shows that the goal of the interaction-based allocation scheme to reduce the number of mixed conflicts was fulfilled, as the ATCOs were mostly occupied with resolving conflicts between flights that were under their manual control.

Given the difficulty of linking gaze to specific flights when flight symbols and labels are overlapping or closely spaced, the results do not provide conclusive evidence that automated flights were less visited in the interaction-based scenario (Hypothesis H3). It did provide proof, however, that automated flights in general evoke less gaze than manual flights and that the interaction-based allocation created more visual segregation between manual and automated flights. This might help prevent the 'gray swan' phenomenon described by Wickens (2009). Here, gray swan refers to an unexpected event that the operator could have anticipated (i.e., the ATCOs were aware that mixed conflicts could occur and that they would require their involvement). An example was discussed in Chapter 3 where one of the ATCOs was surprised by a mixed conflict in which the automated flight was completely surrounded by other automated flights, suggesting that the ATCO did not scan the automated flights well, or not at all. Novel support tools that fade irrelevant flights (e.g., Kumbhar et al., 2024) or highlight conflicting flights (e.g., MUAC's lateral obstacle and resolution display, Eurocontrol, 2024b) on the controller working positions can help further shield ATCOs from overlooking mixed conflicts.

With respect to control resolution efficiency, Hypothesis H4, the interaction-based scenario showed indeed a small advantage with slightly more flights flying shorter routes. The ATCOs themselves also considered this scenario to be more efficient, partly due to the fact that the automation felt more like a teammate, rather than an independent entity interfering with their own work. Together with the increased sense of safety, this was one of the primary reasons that they showed higher appreciation for the interaction-based scenario, meaning that Hypothesis H5 was confirmed.

6.6.2 Automation

The lack of letters-of-agreement adherence in the automated agent led to interactions between Amsterdam inbounds and other flights that are not present in current-day operations. More specifically, automation should have ‘known’ that flights towards Amsterdam should be at FL250, 10 miles *prior* to DENUT (a XCOP near the Dutch border), rather than *at* DENUT. Future research should therefore expand the automation to include at minimum the most important agreements and other sector-specific rules, to strengthen the conformance with ATCO working styles (Westin et al., 2016a). This will prevent the majority of mixed conflicts encountered in the experiment, which should lead to increased ATCO acceptance.

Furthermore, when mixed conflicts do occur, automation should be able to independently solve them if a straightforward solution is available. For example, when the automated flight still needs to climb or descend towards its exit level while the manual flight is already at its exit level, the conflict can be most efficiently solved by adjusting the automated flight. If it cannot be guaranteed that adjusting the automation-controlled flight is the preferred solution, the ATCO should still be able to request the automation to solve specific conflicts. One ATCO suggested that the mixed conflict pairs in the VERA dialog should be color-coded to indicate who has to solve the conflict as also communicated in flight-centric operations (Birkmeier et al., 2011).

6

6.6.3 Experiment

The resemblance of the SectorX simulation environment with respect to the operational system can be increased by several small adjustments, such as correcting mouse inputs and fine-tuning the STCA settings. This will reduce cognitive workload unrelated to the flight allocation or automation and likely increase ATCO acceptance. Interestingly, the results on perceived simulation fidelity do not deviate much from those in Chapter 3, despite extensive development of the simulator. The increased realism and additional features greatly expanded the face validity of the simulator, but at the same time made minor differences with the operational HMI more prominent. Aircraft behavior could benefit from using a probabilistic model, such as developed by Pepper et al. (2023), which should improve climb performance realism.

The use of scenarios based on real traffic samples was a two-sided coin. On the one hand, it meant that the experiment was close to the current operation, assessing a potential ‘quick’ implementation. But on the other hand, the allocations may have had a smaller effect than what would be possible if traffic patterns, routes and sector geometries were optimized for the proposed concept of operations.

As with most human-in-the-loop experiments, training time was a balance between participant availability and required familiarization with the task at hand. According to Balfe et al. (2018), understanding of automation is perhaps even more important in safety-critical systems (rail signaling operators in this case) than the capabilities and reliability of the automation. This was echoed in automated driving experiments by Khastgir et al. (2018) where introducing knowledge about a system and its limitations facilitated high levels of trust regardless the capabilities of the automation. The short briefing and training duration in this chapter's experiment severely limited the amount of understanding that could be attained, despite the simple rule-based form of automation. In an ideal world, each ATCO would have trained for a prolonged time, on multiple days and measurements would have been taken for multiple distinct scenarios. This would not only have increased trust in the automation, but would have also supported a wider applicability of the results beyond the chosen sectors and scenario.

6.7 Conclusions

This chapter confirmed that automating the handling of part of the traffic in an en-route airspace shared with a human ATCO is feasible under certain conditions. A human-in-the-loop simulation exercise showed that minimizing interactions between automated and non-automated flights through a smart allocation scheme, leads to increased ATCO acceptance and a reduction in second-guessing automation capabilities compared to a more pragmatic flow-based scheme. However, the overall efficiency was only marginally better in the interaction-based allocation. To increase the operational applicability and further reduce the occurrence of mixed conflicts, future research should ensure that the automated agent has a better notion of the 'rules' that ATCOs use, such as those set out in letters of agreements. In addition, a form of automation that allows the ATCO to request it to solve a conflict is expected to significantly increase mixed-conflict resolution efficiency.

7

Discussion and conclusions

This chapter revisits each of the chapters of this thesis and reflects upon their findings in relation to the research questions outlined in Chapter 1. Limitations of the study are provided, together with recommendations for future research in the field of human-automation teaming in an air traffic control environment. The chapter concludes with a discussion on the thesis' operational relevance and a glimpse into the future.

Parts of this chapter have been published in:

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

7.1 Introduction

The air traffic control (ATC) domain is expected to undergo a paradigm shift with the introduction of advanced automation, capable of autonomously directing flights through the airspace. Unless all ATC services are completely automated and all (legal) responsibilities lie with the automation designers, human air traffic control officers (ATCOs) will continue to play a central role. The gradual automation of functions, and the subsequent phasing out of human involvement, leads to a large number of fundamental issues, as operators seem to be inevitably pushed into a supervisory role (Bainbridge, 1983, Strauch, 2018). This leads to a serial system where humans monitor a (usually partly) automated system, while performing a myriad number of tasks which could not be automated (yet) (Endsley, 2017).

Many of these issues can be diminished or even avoided if the in-between state with partial automation is skipped by only introducing automation that is mature enough to act autonomously. This approach is referred to as the 'cliff-edge' principle by Young and Stanton (2023). Because the maturing of automation in a safety-critical environment like ATC is no easy feat, adhering to this principle could lead to significant delays in the introduction of higher levels of automation (LOAs).

However, the cliff-edge principle can also be applied in conjunction with a constraint-based automation strategy that limits the high(er) LOA to a subset of the work domain (i.e., for some of the flights in an airspace). The subset can then be expanded in line with technological advances. Limiting the number of (potential) active LOAs and providing a clear separation of responsibilities can help prevent confusion among human operators about the current system state (Endsley, 2017).

A constraint-based automation strategy had not yet been extensively researched in an ATC context, despite its hypothesized benefits. Nevertheless, Eurocontrol's Maastricht Upper Area Control Centre (MUAC) has embraced this strategy in its ongoing ATC Real Groundbreaking Operational System (ARGOS) as the first air navigation service provider (ANSP) (Eurocontrol, 2024b). Many hurdles are yet to be overcome on the way towards full operational implementation, however. In order to address some of these hurdles, this thesis set out to achieve the following goal, as was defined in Section 1.5:

Research goal

Establish how flights can best be distributed between a human ATCO and an automated system, sharing control of an en-route sector, such that interference between the two agents/entities is minimized.

The chapters of this thesis were structured around five research questions, which are discussed in Section 7.2. These included a literature survey and four simulation exercises with professional en-route ATCOs, which lead to new insights on constraint-based strategies and teamwork in general, as reflected upon in Sections 7.3 and 7.4. Next, Section 7.5 discusses the limitations stemming from the research scope and subsequent recommendations for future research. Lastly, Section 7.6 dives deeper into the operational relevance of the findings and summarizes what needs to happen before the researched concept can advance to operational deployment.

7.2 Retrospective

The chart in Figure 7.1 was introduced in Section 2.3.1 to describe and compare a function- and constraint-based automation strategy. To recall, a function-based approach simultaneously increases the LOA for *all* flights in an airspace, while a constraint-based approach allocates certain flights to a high(er) LOA than others and gradually tilts the balance towards the higher LOA. That is, more flights are increasingly handled at the higher LOA. Each chapter of this thesis covers one or more of the elements in the chart:

Chapter 2 contains a literature survey on strategies to raise the LOA (i.e., move from left to right in the figure).

Chapter 3 describes an initial experiment to empirically assess ATCO stances on the concept of allocating flights either to themselves or to automation in a shared airspace (i.e., the allocation of flights to two distinct LOAs as in a constraint-based strategy, moving up and down in the figure).

Chapter 4 continued with a discussion and experiment on the impact that delegating flights has on an ATCO's conflict detection and resolution workflow (i.e., the interdependence of flights at either LOA). The experiment focused on mixed conflicts in a worst-case scenario, where ATCOs would have completely overlooked automated flights that may interact with flights under their control.

Chapter 5 aimed to find a method to determine the complexity of individual flights, which can then be used to automatically allocate flights to the low or high LOA.

Chapter 6 combines all insights from the preceding chapters to empirically assess an interaction-based flight allocation scheme, inspired by Chapter 5.

7

This section discusses each research question, grouped according to the three phases of the research: 1) exploration, 2) model, and 3) validation. After stating the research question, an itemized overview of the most important conclusions is given, followed by a more elaborate discussion on their implications and relations.

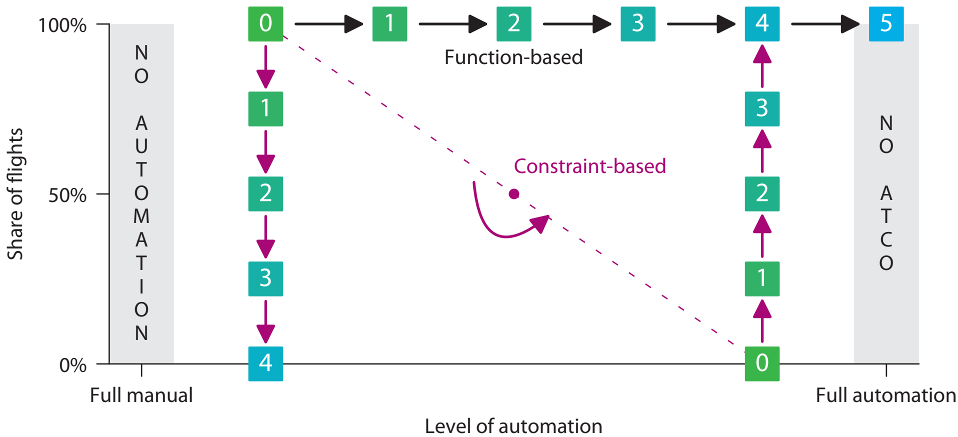


Figure 7.1: Level of automation chart, illustrating the function- and constraint-based automation strategies introduced in Section 2.3.1.

7.2.1 Exploration

The first phase of this research took an exploratory approach. It focused on investigating automation strategies reported in literature and was followed by an initial human-in-the-loop experiment to assess the proposed concept of operations and gauge its feasibility from an ATCO perspective in terms of acceptance, trust, etc. The literature survey was driven by the first research question:

Research question 1 **Chapter 2**

How does the human-automation allocation of flights fit within existing strategies towards full automation in the ATC domain?

main findings

- Most automation projects follow a *function*-based allocation strategy, which evokes a serial human-automation system that ultimately push the human into a problematic supervisory role (leading to transient workload peaks, skill erosion, out-of-the-loop situation awareness, etc.).
- Constraint-based (i.e., *flight*-based) allocation evokes a parallel human-automation system that keeps humans as long as possible in the loop, potentially leading to less skill erosion, better situation awareness, higher job satisfaction, etc.

The vast majority of automation projects in- and outside the ATC domain follow a technology-centered, function-based allocation strategy, where automation increasingly replaces tasks that a human used to perform. Operators are pushed into a supervisory role, leading to a serial system where they monitor a (usually partly) automated system, while performing a myriad number of 'remaining' tasks which could not be automated (yet). The inevitable human-automation issues that come with this have been identified decades ago (Bainbridge, 1983), are still valid today (Strauch, 2018), and are expected to remain in the artificial intelligence (AI) minded future (Endsley, 2023).

A promising way to avoid (many of) these issues is a constraint-based strategy: first attain a high level of automation in a constrained environment, before expanding its application. MUAC is currently pursuing a strategy along those lines in its ARGOS project, in which basic, routine traffic is automated, while the ATCOs remain responsible for the more complex, non-basic, traffic (Eurocontrol, 2024b). Because a constraint-based strategy has not yet been extensively researched in an ATC context, an initial exploratory experiment was performed and discussed in Chapter 3 to answer the second research question.

Research question 2 **Chapter 3**

To what extent is the transfer of control of flights to an automated system dependent on system-proposed allocations, individual ATCO preferences and automation capabilities?

main findings

- ATCOs are willing to allocate control over specific flights to automation, provided that the automation is sufficiently capable and the flight is routine.
- Flight allocation schemes should not be (solely) based on geographic areas or a flight's total vertical movement within the sector.
- Limited automation capabilities prevent ATCOs from fully delegating flights.
- Experiencing automation first-hand is important for ATCO trust and acceptance, advocating 'innovation through simulation'.

Six professional ATCOs were subjected to six different flight allocation schemes based on either a flight's location, a flight's vertical change within the sector, or on an 'all-or-nothing' concept with either all flights allocated to the ATCO or to the automation. The ATCOs were given total freedom in deviating from the proposed strategies, which most of them took to heart, complicating the comparison of these strategies. Later experiments have restricted the allocation freedom, because no conclusive in-between participant (or suggested allocation) comparisons could be drawn from this experiment. Nevertheless, it did result in some first insights regarding the desired allocation schemes.

In their reflections, ATCOs appeared to avoid potential interactions between human- and automation-directed flights. ATCOs preferred to keep flights requiring (large) flight level changes with them, although this was partly caused by the limited capabilities of the automation (e.g., it could not issue direct-to clearances). These results suggest that interactions between flights (should) play an important role in developing flight allocation schemes from the perspective of operator acceptance.

After some initial skepticism following a short training session, the ATCOs reported a great deal of trust in the automated agent. First-hand experience was essential for this, highlighting the importance of empirical research by real-time, human-in-the-loop simulation exercises. The simple rule-based solver helped in creating a predictable 'colleague', although the ATCOs would have appreciated additional communication options with that 'colleague', especially regarding planned top of descent locations.

The experiment roughly corresponded to Level 2 of the ARGOS taxonomy discussed in Section 2.5.1, where ARGOS only (but fully) controls individual flights allocated to it and the ATCO controls all other flights. However, in MUAC's vision the minimum LOA of the manual flights would be higher than in the experiment, with ARGOS suggesting plans for all flights. In this regard, MUAC envisions a more serial system at this level, as the ATCO needs to check the suggested plan which contradicts the cliff-edge principle. Instead of management by consent or exception (where the automation proposes a solution, Billings, 1997), increased decision-support tools that allow ATCOs to make informed decisions, such as MUAC's lateral obstacle and resolution display (LORD, Eurocontrol, 2024b), may be more appropriate for complex flights. In a human-centered system it is sometimes better to maintain a lower LOA, even when technologies that can provide a higher LOA are available (Kaber and Endsley, 2004).

7.2.2 Flight allocation model

The second research phase aimed to dive deeper into the conflict detection and resolution (CD&R) task of ATCOs and the intrinsic complexities of flight-based control allocation. With the goal to obtain empirical quantification for both cognitive effort and flight complexity, this phase's two chapters would provide valuable input for a better allocation scheme than the simple and pragmatic schemes suggested to the ATCOs in Chapter 3.

Chapter 4 set out to take a closer look at the CD&R task of an ATCO and how their work flows change when an automated agent is introduced in the workspace. It focused on analyzing CD&R tasks, because these are the most fundamental to an ATCO's objective of managing safe and efficient air traffic.

Research question 3 **Chapter 4**

To what extent is the workflow of ATCOs affected by flights delegated to an automated system that interact with flights under their responsibility?

main findings

- Conflict detection by ATCOs cannot be captured in a serial, step-wise process that can be readily quantified in terms of cognitive effort by measuring its duration.
- Conflict detection time is mostly determined by the number of new, unseen flights, rather than the total number of flights in the sector.

Numerous previous studies, combined with new observations, led to the creation of flowcharts that mimicked the likely thought processes of ATCOs in a structural way. In general, ATCOs initially filter flights based on their vertical separation, followed by their directional overlap and then their temporal overlap. Processing these steps is hypothesized to take increasing cognitive effort, as skill-based behavior is traded for rule-based behavior requiring a prediction of future aircraft states. This is done based on experience and training, but also a rigorous airspace design and the use of extensive letters of agreement (e.g., even-odd flight levels for opposite streams of traffic) help ATCOs in this task.

In case the flights in an airspace are shared between an ATCO and automation, mixed conflicts can occur where the ATCO has not been actively involved with a flight that interacts with their flights. In the worst case, the ATCO has not seen an automated 'blue' flight at all, until it suddenly poses a problem to one of the flights under their manual control. An example of this was observed in Chapter 3, where an ATCO did not spot a mixed conflict because the automated flight was emerging from *"a sea of blue aircraft"*. To simulate this worst-case scenario in an experiment, ten ATCOs first performed CD&R on simplified static scenarios with only 'green' flights (i.e., under manual ATCO control). Then in a second phase, one or two blue flights would pop up in each scenario, requiring the ATCO to re-evaluate the conflict status, and in some cases also adjust previously issued clearances.

Despite the theoretical and observational foundation, quantifying the pop-up flight's impact on the CD&R workflows turned out to be challenging. Human ATCOs, or experts in general, do not appear to follow strictly serial thought processes when performing tasks (Rasmussen, 1986). Even in a simplified and focused environment as simulated in the experiment, mental shortcuts and frequent task switching or parallelism seem to occur. While flowcharts are an appropriate means for analyzing, structuring and communicating general thought processes, they are inevitably a simplification and are not meant to capture the entire cognitive process of ATCOs. Perhaps the aim should not be to quantify these processes in numbers but rather to qualitatively describe these processes based on less isolated, more dynamic and realistic scenarios. Observing a large number of such scenarios may provide a sufficiently sound impression of which type of situations are cognitively most demanding.

The flowcharts showed to be useful later on in the investigation, namely in the design of the automation solvers used in the experiments from Chapter 3 and Chapter 6 (see Appendix A.5 for details and pseudo-code). They were also used by Kumbhar et al. (2024) in a flight-filtering algorithm that fades irrelevant flights for an ATCO based on their level of interaction with a selected flight (which was validated on the results from Chapter 5). Continuing along this line, they may be useful for substantiating the complexity of individual flights and/or adapting the automated handling of flights, to better match human-like reasoning and with that increase ATCO acceptance.

To obtain a baseline measure for perceived individual flight complexity and answer the fourth research question, 15 ATCOs from three different MUAC sector groups were asked to rate the complexity of 36 flights overlaid on a static radar snapshot (Chapter 5). In comparison to Chapter 4, the use of actual traffic snapshots lead to more realistic and complex traffic scenarios. Previous studies focused on the complexity of an entire traffic scenario, rather than individual flights, while assessing the complexity of every single flight is a prerequisite for the automated allocation of flights proposed/investigated here.

Research question 4 **Chapter 5**

Which other flights in the airspace add to the perceived complexity of an individual flight and what characterizes them?

main findings

- ATCOs are largely consistent on which flights do and do not add to the perceived complexity of another flight.
- Interactions between flights (i.e., closest spacing and reduced solution spaces) play a major role.

Personal differences set aside, the ATCOs in the experiment largely agreed on which flights were contributing to the complexity of a single flight of interest. This finding paves the way for a flight allocation algorithm which does not need to be tailored for personal preferences. ATCOs can then expect, and work with, predictable automated decisions, which helps buildup and maintain their trust in the system. From a legal point of view, this is preferred as well, as it is easier to certify a single static configuration algorithm. ATCOs can then fine-tune the allocation when desired if they are allowed to take manual control over any flight they want.

Results indicate that (potential) interactions between flights are a primary factor adding to complexity as perceived by ATCOs. Flights that overlap in altitude *and* get within 10 NM of each other were overrepresented among the flights that the ATCOs selected as contributing to the complexity. Nevertheless, also flights that did not meet these criteria were included, suggesting that a flight allocation algorithm should take more parameters into account. For example, there could be an interdependence between the complexities of flights that was not tested in this thesis. In other words, is there a carryover effect that raises the complexity of an otherwise low-complexity flight, when it interacts with a high-complexity flight?

Another remaining question is whether flight-centric complexity alone is sufficient to determine flight allocation. The ATCOS indicated a preference to delegate low-complexity flights to an automated agent in both Chapters 3 and 6. Although this is in line with MUAC's proposed strategy to automate basic flights first (Eurocontrol, 2024b), it did not involve the full work domain of an ATCO.

Irrespective of these potentially missing links, an initial flight allocation algorithm does not need to be perfect, as long as it only allocates flights to automation that are guaranteed to be of low complexity. It is operationally preferred to, perhaps inadvertently, allocate low-complexity flights to the ATCO than it is to allocate high-complexity flights to the automation.

7.2.3 **Validation**

With the final experiment, Chapter 6 consolidated all lessons learned and involved a realistic dynamic traffic scenario with fixed allocation schemes to address the pitfalls from the freedom given to the ATCOs in Chapter 3. It served to validate the idea originating in the preceding chapters that a constraint-based flight allocation works best when interactions between flights are considered. The last research question was therefore:

7

Research question 5 **Chapter 6**

Given a realistic traffic scenario, how should flights be allocated to either the human or automation, such that interactions between human- and automation-controlled flights are reduced, combined team performance is best supported and ATCO acceptance is increased?

main findings

- Flights having (close) interactions should be assigned to the same agent, to increase ATCO acceptance and reduce second-guessing of the automation.
- Both agents should follow the same rules and standards, adhering to established letters of agreement.
- The automation should timely inform ATCOs of mixed conflicts, especially those that it will not or cannot resolve.

Because this thesis did not investigate the desired complexity threshold at which flights should be allocated to automation, a fixed 50% distribution of the flights to either agent was chosen. In light of the findings in Chapter 3 this experimental restriction proved to be essential for a good comparison between the two designed allocation schemes. Nevertheless, almost all of the ATCOs involved in the experiments stated that they should always be able to take over *any* flight from the system, if this were to be implemented in operations. Whenever ATCOs are held accountable for something, they should be able to intervene as stressed by themselves (Bekier et al., 2012) and human-factors experts. In general, operators of highly automated systems should always be able to establish 'meaningful control', regardless of whether they are ATCOs, pilots or other operators working on the sharp-end of complex sociotechnical systems (Holford, 2020).

The interaction-based allocation scheme, which led to a reduction in mixed conflicts, was indeed preferred by the ATCOs over a more pragmatic flow-based scheme. Visual attention for flights allocated to the automation reduced when these flights interacted less with ATCO-controlled flights. In addition, second-guessing automation decisions and actions was reduced and the overall efficiency in terms of track miles was slightly higher. A flow-based scheme may benefit from human and automation both following the same procedures and rules from letters of agreement, which was lacking in the experiment. However, because such procedures will not preclude all interactions, an interaction-based scheme is still encouraged.

7.3 Reflections on a constraint-based approach

This thesis only considered a purely constraint-based automation strategy, with flights either at a very low or a very high LOA. However, flights at the low LOA could also benefit from increased automation support beyond the basic level used in the experiments, especially considering that these flights are by definition the most complex cases. Combining the two strategies fosters faster innovation than a pure function-based approach, but still acknowledges the incremental nature of technological advances and step-wise changes in policies.

It is thus not surprisingly that MUAC's ARGOS strategy also involves a combination of a function- and constraint-based approach (Figure 7.2). The question is then what should define the maximum LOA at which flights can still be operated manually, before the human-automation issues that are best avoided become apparent. Consolidating everything so far, the information acquisition and analysis stages can be largely automated for these flights, without negative effects on human performance. The execution of actions can also be automated to a great extent, provided that the human takes the initiative and cannot be surprised by silent action execution (i.e., management by exception). However, it is fundamental that the decision-making process should remain with the human operator. Support for this can be provided by the automation, but should not transcend the information stages (i.e., the input). Examples of this can be found in MUAC's LORD (Eurocontrol, 2024b) and solution space diagrams (Klomp et al., 2019), that both show the ATCO which clearances would lead to or resolve a conflict with other traffic, rather than selecting and proposing a single solution.

As discussed before, the allocation of flights in a constraint-based system to a specific LOA should be based on each flight's complexity, with lower complexity flights assigned to the higher LOA first. These complexities are not necessarily constant throughout a flight's traversal through the sector though. Chapter 3 already showed that ATCOs frequently delegated flights after passing the 'challenging' part of their route, i.e., the climb,

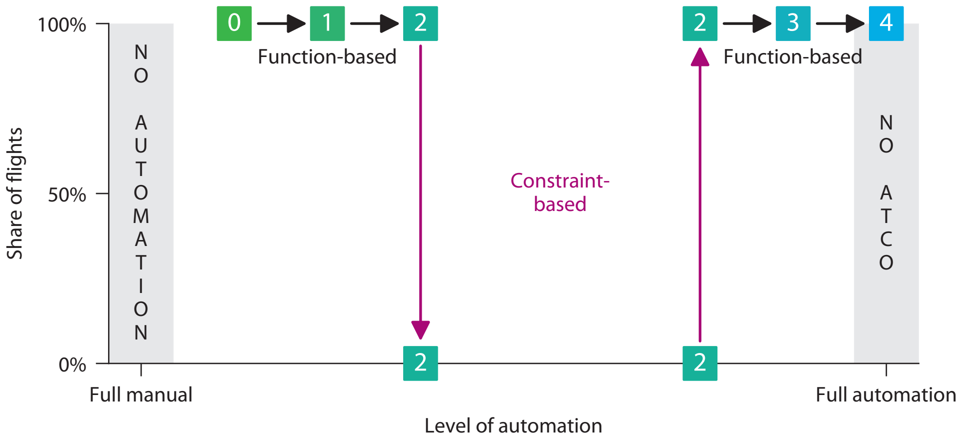


Figure 7.2: Level of automation chart for MUAC's ARGOS, as introduced in Section 2.5.1.

descent or conflict situation. External factors, such as adverse weather conditions and associated reroutes, can also play a role as these affect trajectory uncertainty (Corver and Grote, 2016). Perhaps flights should thus be re-allocated when their complexity changes beyond a certain threshold and basic flights become non-basic, or vice versa. Especially when sectors become larger, a significant portion of the flight trajectories may be basic as the cruise part of flights will make up a relatively large part of their time in the sector. The predicted growth in air traffic (density) from the 10.1 million flights that transitioned the European airspace in 2023 to 12.2 million flights (+21%, or +10% from pre-COVID-19 numbers) in 2030 (Eurocontrol, 2024a) can invalidate this. However, smarter and perhaps also even dynamic sector designs may succeed in distributing flights more evenly, which can lead to a decrease in interactions between flights and damp the trend.

Furthermore, the threshold at which flights are classified as basic or non-basic should not only be complexity-dependent but also depend on human factors. To maintain all the benefits of constraint-based automation allocation, the number of flights allocated to the ATCO should not drop below a certain base line. As seen in Chapter 3, when ATCOs delegate (almost) all flights to automation, they are at risk of becoming bored or unskilled (Kenny and Li, 2022). A consequence of enforcing a minimum number of manual flights might be that flights that would normally be considered basic are assigned to the ATCO, and not much would be gained. Another option would be to dynamically enlarge or adjust the airspace assigned to the ATCO, but this may clash with other considerations such as the license/training of the ATCO. The total number of flights should be capped to limit the potential of simultaneously occurring non-standard events (e.g., emergencies, pilot requests) that require ATCO attention.

The threshold may be dynamically adjusted based on human performance through physiological measures (e.g., brain activity, Borghini et al., 2017) or control performance (IJtsma et al., 2022). This could also cater for the (temporary) offloading of additional flights when the ATCO has to deal with extra complex flights.

7

7.4 Reflections on teamwork

Although teamwork was not an explicit research objective in this thesis, the results and insights collected throughout the experiments do lead to a number of observations. The overarching question in this section is whether constraint-based automation can lead to more human-automation teamwork in ATC.

In the experiments of Chapters 3 and 6, both agents worked towards the same goal: ensuring safe and expeditious traffic flows within the sector. The interdependence of each team member's tasks in achieving this goal is a crucial aspect in teamwork (O'Neill et al., 2023). As discussed in Section 1.4, there has to be a certain overlap between team members in order to establish a sense of teamwork, but in our example of ATC this overlap should *not* lead to automation and ATCO thwarting each other in trying to resolve many mixed conflicts. Much effort in this thesis was therefore put in creating a parallel system where the ATCO and the automation interfered as little as possible (e.g., by using an interaction-based allocation scheme in Chapter 6).

Only one of the six ATCOs in Chapter 3 *perceived* the experiment's setup as teamwork. The more parallel interaction-based allocation in Chapter 6 received more favorable teamwork opinions, but both experiments stayed far from genuine teamwork according to the ATCOs. In light of the Big Five model by Salas et al. (2005), several important aspects defining teamwork were indeed missing, such as team leadership and orientation, and closed-loop communication. If the automation was programmed to operate more like the ATCO and follow the same standard procedures (i.e., letters of agreement), their work may require less segregation because interference will be reduced. On the other hand, limited mutual performance monitoring was established (e.g., warning ATCOs through VERA when a mixed conflict was unresolved).

7

In this regard, one might question whether the allocation of flights to *either* an automated system *or* an ATCO should be considered human-automation teamwork (HAT) at all. The tested form is perhaps closer to human-automation interaction (HAI, Janssen et al., 2019) or partnership. According to O'Neill et al. (2023), the LOA applied to an agent determines whether discussing HAI or HAT concepts is more appropriate for a system under consideration. In the constraint-based setup researched in this thesis, the agents are operating at two different LOAs, so there is not a single system-wide LOA. For flights that are completely allocated to the automation, HAT might be the better fit, whereas a HAI perspective is more fitting for the manual flights with limited automation support.

Tentatively, it may be more important to establish teamwork in serial systems than in parallel systems. With the ATCO responsible for non-basic flights, the majority of human-automation interactions will take place between basic and non-basic flights. Instead of providing automation-generated proposals to the ATCO, it might be more beneficial to provide the human ATCO with alternatives after they have come up with their own solution. These alternatives should then follow the same constraints and goals that the ATCO uses. For example, if the ATCO is about to turn a flight 'left' to establish a certain separation, the automation might show that turning 'right' would establish the same separation but with less delay for the flight. In this case, a certain 'improvement' threshold should be established, to prevent the automation from annoying the ATCO with micro-optimized alternatives that provide no significant benefits. A system like this reverses the roles in the learning-on-the-job scenario brought up in Section 2.3.2, with the automation now

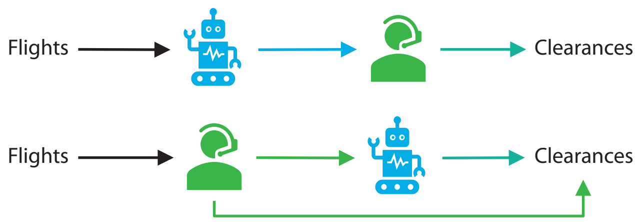


Figure 7.3: Two forms of serial automation, with the human backing-up the automation (top) or vice versa (bottom).

becoming the supervisor (or mentor) of the ATCO, as shown in Figure 7.3. Of course it is then important to prevent the ATCO from relying solely on the automation identifying human errors (i.e., automation misuse). In addition, the ATCO should be able to bypass the automation, e.g., in case of a system failure.

7.5 Limitations and recommendations

This thesis had to scope down the research conditions as outlined in Section 1.6 to meet time and resource constraints. This inevitably affected the generalizability of the results and their direct applicability to operational ATC environments. On the other hand, some of the limitations found in previous research were avoided, broadening the applicability of the results. For instance, the use of professional ATCOs and a medium-fidelity simulator that mimicked their actual working positions led to significantly higher face validity compared to many other experiments that used novices or greatly simplified generic simulators. Nevertheless, there are numerous recommendations for future research, which are discussed in this section together with the implications of the chosen scope. Each part ends with some take-away recommendations.

7.5.1 Experimental evaluation

Despite the medium-fidelity simulator and professional ATCO participants, experiments were still simplified with respect to actual ATC operations. There was no weather (wind, turbulence etc.) and there were no emergency flights or pilot requests that would deviate from the flight plans. The unpredictability of real-life traffic is an important reason why we cannot increase the LOA. Humans and especially trained operators are naturally very good at adapting to changing situations while this has proven to be much more difficult for automation. In the perfect world of a simulation, automation can cope with any situation that is thrown at it. In reality, however, there will always be unexpected situations that were not foreseen and of which it is unknown a-priori how the automation would need to act (i.e., an open system). That is, unless the automation is limited to rule-based behavior, a topic that is further discussed in Section 7.5.2.

The absence of voice communication was a major contributor to the relatively low workload reported in the experiments from Chapters 3 and 6. Radio transmissions are indeed a major element of the current ATCO work and an important focus of simulation fidelity in ATCO training (Dow and Histon, 2015). However, it is not unrealistic to foresee that data links will replace many instructions that are currently still transmitted via voice.

For example, 'assume' and 'transfer' of flights are occasionally already performed via controller-pilot data link communications (CPDLC) at MUAC. For short-term time-critical instructions, voice is expected to remain the prevalent means of communication due to its intricate speed and instant feedback/acknowledgment. However, it is precisely these flights that will stay with the ATCO in a shared human-automation airspace. It is worth noting that the shift from radio transmissions to data link has been shown to not only lower workload but also decrease engagement (Martins et al., 2024).

In contrast to many experiments on en-route conflict detection and resolution, this thesis was not limited to the horizontal plane and included the altitude dimension. The vertical traversal of aircraft is a key complexity factor that cannot be neglected as echoed by the involved ATCOs and the results of all experiments. In contrast, the assumed redundancy of speed control in en-route operations turned out to be somewhat limiting the ATCOs in finding (efficient) solutions and complicating the resolution task. This was especially the case with overtake conflicts, as often seen in sectors prone to bunching such as the western part of the Brussels sector. Future experiments should open up the full control-domain to ATCOs.

Despite numerous simulator improvements between Chapter 3 and Chapter 6, the ATCOs' subjective simulator fidelity ratings did not improve. This striking result shows the irony that comes with increased realism and immersion: small inaccuracies become more pronounced when the environment is more immersive. That is, the fidelity of the lowest-fidelity component may have an excessive impact on a simulation's total fidelity (Schricker et al., 2001). Nevertheless, the use of a simulator with high face validity made it easier for the ATCOs to appreciate the proposed concept as a viable system, rather than a vague vision for the far future. It also encouraged the development of dual-use for existing tools and displays for automation-related functions, such as the automation highlighting mixed conflicts with VERA.

Ideally, longitudinal studies are performed in which ATCOs receive multiple days of training before going through a number of trials involving different scenarios and/or setups. Because some of the ironies of automation only emerge after a considerable amount of time has passed (e.g., boredom) the relatively short studies in this thesis may have underreported the benefits of flight-based control allocation. As the concept transcends from its initial stage, such studies may become more feasible.

main recommendations

- Include uncertainties such as weather, pilot requests or system failures.
- Expand the control dimensions to the full scope available to ATCOs (e.g., speed instructions and vertical rates).
- Run longitudinal studies with extensive training to let the ATCOs get accustomed to the automation.

7.5.2 Automation control capabilities

In Chapter 3, the ATCOs frequently took manual control solely to issue direct-to's, which the automation could not do. Capabilities of an automated agent are therefore perhaps more important in flight allocation than the ATCOs themselves may realize (Figure 3.17). The automation in Chapter 6 was upgraded; the agent could now issue direct-to's and assume and transfer flights near the sector borders. Although the ATCOs could not allocate flights themselves in the experiment – and were thus forced to let the automation handle situations, or to work around automation-directed flights – subjective feedback indicated a heightened appreciation for the automation with increased capabilities.

Nevertheless, mixed conflicts occurred in both chapters that could have been prevented if the automation would have adhered to the rules from established letters of agreement that the ATCOs did follow. Making these agreements, that are often published in unstructured documents, machine-readable has been accelerated by the recent advent of natural language processing techniques (Batra et al., 2024). Doing so will equalize the constraints that human and automation use in handling air traffic and help establish a common ground. However, present agreements may need to be revised to suit the introduction of an automated agent.

An important assumption in this thesis was that the automation is infallible, i.e., ATCOs did not need to monitor its proper functioning. In fact, the automation should self-monitor and alert the ATCO when it detects it is misbehaving, or if it encounters situations that are outside its constraints. In the ARGOS project, an independent autonomous monitoring system is envisioned to take this role (Lanzi et al., 2021). When failures do occur, they should be 'graceful' and not remove all automation support at once (Edwards and Lee, 2017). Even if the ground-based automation were infallible, the air-side (i.e., pilots or aircraft/autopilot) may fail to respond to or even deviate from issued clearances. One such case was inadvertently encountered by all ATCOs in the flow-based scenario of Chapter 6 when an automated flight did not execute an instructed climb, which triggered a yellow 'LVL' text alert in the flight's label. Although the ATCOs' responses to this event were not explicitly measured, anecdotal evidence indicates that most ATCOs were surprised by this occurrence and did not immediately understand whether it was a fault of the automation, the pilot or the simulator. Bringing the ATCO back in-the-loop in such situations is a research topic on its own, but may benefit from the constraint-based approach, where ATCOs are actively involved with (part of) the traffic, rather than purely monitoring.

main recommendations

- Equip the automation to perform all ATCO tasks, including the resolution of mixed conflicts for which flights under its control can provide a logical and straightforward solution.
- Feed the automation with letters of agreement to bring its clearances and their timing more in line with ATCO behavior and expectations, which should minimize the occurrence of mixed conflicts.

7.5.3 Feedback and communication between human and automation

The automation was designed to be 'simple' and rule-based, with the intention to eliminate the need for extensive communication and long familiarization sessions to build sufficient trust and proficiency with the ATCOs who had to work with it. This assumption turned out to be overly optimistic. As articulated by Christoffersen and Woods (2002), more automation generally requires more communication between human and machine, not less. Lack of communication is a common pitfall in automation design that hinders the establishment of productive human-automation teamwork (Norman, 1990).

Consensus within the cognitive engineering and AI communities also points to the requirement that automation should disclose information on its capabilities, limitations, what task(s) it is currently doing, why and how it is doing the task(s) in the specific way(s) that it is, and what it plans to do next. This type of feedback is commonly referred to as 'seeing-into' transparency (Chen et al., 2020, Jamieson et al., 2022). However, opening the 'black box' may also come with new challenges related to clutter and the complexity of (visual) representations, potentially leading to increased workload and delayed responses (Van de Merwe et al., 2024, Springer and Whittaker, 2020).

Especially for ATCOs, who prefer to have a clean and uncluttered radar screen, 'minimalistic' feedback solutions are sought that, for example, can communicate machine intentions via decision-support tools (e.g., VERA) that ATCOs are already using today. The ATCOs participating in the experiments of Chapters 3 and 6 reported that they, in particular, missed transparency on the automation's planned top of descent. MUAC's recently introduced display of extended projected profile points from the aircraft's flight management system demonstrates a potential visualization that could be re-used for this purpose (Jagasits, 2024).

In an initial effort to enhance communication that supports the ATCO, Chapter 6's automation alerted the ATCO when it had identified a mixed conflict that it would not solve. The conflicting pair was automatically added to VERA. This was indeed well received by the participating ATCOs, as it corresponds to current practice, where the coordinating ATCO, or ATCOs from adjacent sectors, can 'flag' imminent conflicts for the executive ATCO in a similar way. A next step would be to declutter the screen and help ATCOs focus by fading-out automation-directed flights that are not relevant to any of the human-directed flights (Finck et al., 2023a, Kumbhar et al., 2024). This is especially important as ATCOs will be working busier and/or larger sectors when part of the traffic is automated.

Interestingly, several empirical studies in ATC have reported limited benefits of automation transparency. For example, Westin et al. (2022) showed that ATCOs' acceptance of machine-generated resolution advisories was more affected by matching them to human preferences and strategies (i.e., conformance) than by offering increased transparency. This suggests that understanding and accepting machine intentions can also be achieved by matching the automation's behavior to ATCO best practices, preferences, and expectations (Westin et al., 2016b). In addition, some form of 'letter of agreement' between human and digital ATCOs can further reduce the need for inter-agent communication and coordination, similar to how standard instrument departures and terminal arrival routes minimize the need for communication between ATCOs and pilots.

Eventually, when automation is sufficiently reliable and has proven itself in a variety of challenging circumstances, the need for 'seeing-into' transparency may diminish altogether. This is no different from high-performing human teams, where team members do not need to understand and communicate each other's intentions as long as the work is done well and team members can rely on and trust each other.

While the preceding discussion focuses on communication *from* the automation to the ATCO, communication in the other direction should not be ignored. Following comments by many of the participating ATCOs, they should be able to 'nudge' the automation, just like they can with their human colleagues (Klein et al., 2004). For example, ATCOs need to be able to exclude certain flights from being taken by the automation, e.g., to prevent the automation from complicating the ATCO's sector plan. However, this does come with several risks, such as unclear responsibilities, and additional workload. When the current ATCO-dyad is maintained, the coordinating controller can take this 'managerial' role. However, in future single-controller operations (Gerdes et al., 2022) the sole ATCO should be able to perform this task.

main recommendations

- Allow ATCOs to take ownership over specific mixed conflicts, such that the automation will not try to resolve these.
- The automation should adhere to letters of agreement to minimize communication needs.

7.6 Operational relevance, future outlook

New concepts take time to transition from the experimental phase to full operation. Back in 1987, Hunt and Zellweger (p. 19) already claimed that *"the FAA's new technology ATC computer system will lead to highly automated ATC by the turn of the century."* And yet here we are in 2025, still talking about highly automated ATC as a future goal. As another example, flight-centric ATC was first proposed over two decades ago (Duong et al., 2001) and is still not in operational use anywhere. Giving full decision-making authority over flights to an automated agent is not something for the short term. However, it does seem plausible that specific flights will be handled at a high(er) LOA in the coming decade(s).

The Single European Sky Air Traffic Management (ATM) Research (SESAR) Master Plan also keeps evolving over the years. Its latest edition (SESAR Joint Undertaking, 2024), published shortly before finalizing this thesis, supports the approach taken in this thesis. For the first time, this edition explicitly acknowledges the merits of a form of automation operating within a confined scope. Although this does not necessarily equate to a constraint-based strategy, the trend may continue.

As recently demonstrated by MUAC's LORD (Eurocontrol, 2024b), advanced novel concepts, originating from academic studies like this thesis, can be integrated into existing ATC work environments. A major contributor to that success is the empirical research preceding said integration. Continuing that line, this thesis is a step forward from the many theoretical plans and concepts that have been published over the years. Nevertheless, on its way towards becoming operational, human-automation flight allocation requires additional experiments that go beyond the realism level from Chapter 6, addressing the limitations discussed in Section 7.5.1. The developed software platform, TU Delft's SectorX, provides a solid basis for such further research.

The experiments in this thesis were performed using currently trained ATCOs, on currently operational sectors and traffic. When new automation concepts are introduced, a holistic approach needs to be taken that also looks at these aspects and potentially adjusts them to further optimize combined human-automation performance. ATCOs currently receive extensive training for a particular sector and may need even more training when they become responsible for larger sectors (Klünker et al., 2023), whose design may also change with different routing structures to further optimize the allocation of flights. An example of an innovative concept in that aspect involves flight-centric ATC in combination with moving sectors (Schultz et al., 2023), where 'grouped' flights are assigned to a single ATCO who then handles conflicts within the group as well as conflicts with adjacent groups. In addition, ATCO selection processes may need to be revised in accordance with changes in required skill-sets due to responsibility adjustments (Griffiths et al., 2024).

All this comes with substantial legal implications for ATCOs, ANSPs and system manufacturers (Lanzi et al., 2021). It should be prevented that ATCOs are blamed for following the automation if the automation fails and similarly blamed for not following the automation if they make a mistake. Additionally, who is responsible for automation failures in the first place? As long as such legal barriers have not been fully addressed, transitioning to higher LOAs where automation can take action might not be wise, or even impossible.

7.7 Final reflections

At the start of this thesis, the goal was set to expand the knowledge about allocating individual flights to an automated agent at high LOA, while keeping (large) parts of the traffic in an airspace under manual control. The results of human-in-the-loop simulation experiments indicate that this configuration is indeed viable and can be appreciated by ATCOs, provided that several conditions are met. Most importantly, the automation should be capable of autonomously performing all standard tasks of a human ATCO, should follow the same rules as its human colleagues and should foster a form of two-way communication that is integrated in the controller working position.

To conclude, future airspaces should be organized as shared space, where humans and automation independently and dependently co-operate to facilitate safe and efficient air travel. ANSPs and researchers are encouraged to further research flight-based control allocation and lift the experimental research to a higher technology readiness level (TRL). However, even if new forms and levels of automation can reach a high TRL, this does not mean that they should always be deployed. In a human-centered approach, the human readiness level (HRL) should be the decisive factor in whether a technology is to be used (See, 2021). Now that the ATC domain has celebrated its first centennial and with the imminent advent of advanced automation concepts, ATC should become even more human-centered, not less.

Appendices

A

SectorX ATC simulator

Throughout the experiments in this thesis, SectorX has been used to simulate the work environment of air traffic control officers. For this thesis, the simulator has been vastly enhanced to closely mimic the controller working positions of Maastricht Upper Area Control Centre (MUAC), including commonly used support tools. This chapter provides a brief overview of the various functionalities and implementation details.

A.1 Introduction

The experiments in this thesis made extensive use of SectorX, a Java-based air traffic control (ATC) simulation package developed at Delft University of Technology that focuses on the evaluation of human-machine interfaces (HMI) in human-in-the-loop experiments. The predecessor of SectorX was developed for the MUFASA project on conformal solutions (Westin, 2017), which ran in two editions between 2011 and 2015. Since then it was used for various other research projects. In recent years, SectorX has received numerous enhancements to make it into the full-fledged simulator that it is today. Many of these enhancements were directly driven by this thesis.

A.2 Program structure

A.2.1 Modes

SectorX consists of three modes, all consolidated in a single program pertaining to three standard phases of an experiment:

Editor used to create and preview scenarios. Here sector boundaries, waypoints, routes and flights can be added, deleted or modified on an interactive map. The editor doubles as an operational HMI such that changes in the scenario and their effect on any support tools can be instantly evaluated.

Simulator used to perform an experiment and log data. Using a configuration file, settings can be tuned per experiment run, for example, to enable or disable specific tools.

Viewer used to replay a logged experiment run, replicating the actual interface shown during the experiment by reloading the window setup and configuration parameters. The replay is interactive, making it possible to see what an experiment participant would see and be able to do at any moment in the run (e.g., inspecting solution space diagrams or probing alternative resolution options).

A.3 Human-machine interface

As face validity is an important factor when researching interfaces for professionals in safety critical operations (Dow and Histon, 2014), SectorX is capable of mimicking operational HMIs from various air navigation service providers. Apart from simulating the Maastricht Upper Area Control Centre (MUAC) interface used in this thesis (Figure A.1), SectorX provides support for additional 'styles'. As of yet, LVNL-style positions for area control, approach and tower have been implemented (Figures A.2 and A.3). Providing a high face validity was found to be paramount in getting ATCOs along and increasing acceptance of future tools. By selectively enabling tools and display items, the simulator can be tailored for use by novices or expert ATCOs, minimizing the training duration for both.

The following sections provide a more detailed view of the various tools and elements in the MUAC interface that were relevant for this thesis' experiments. As such, the functionality described here matches the experiments and may deviate slightly from the actual MUAC HMI.

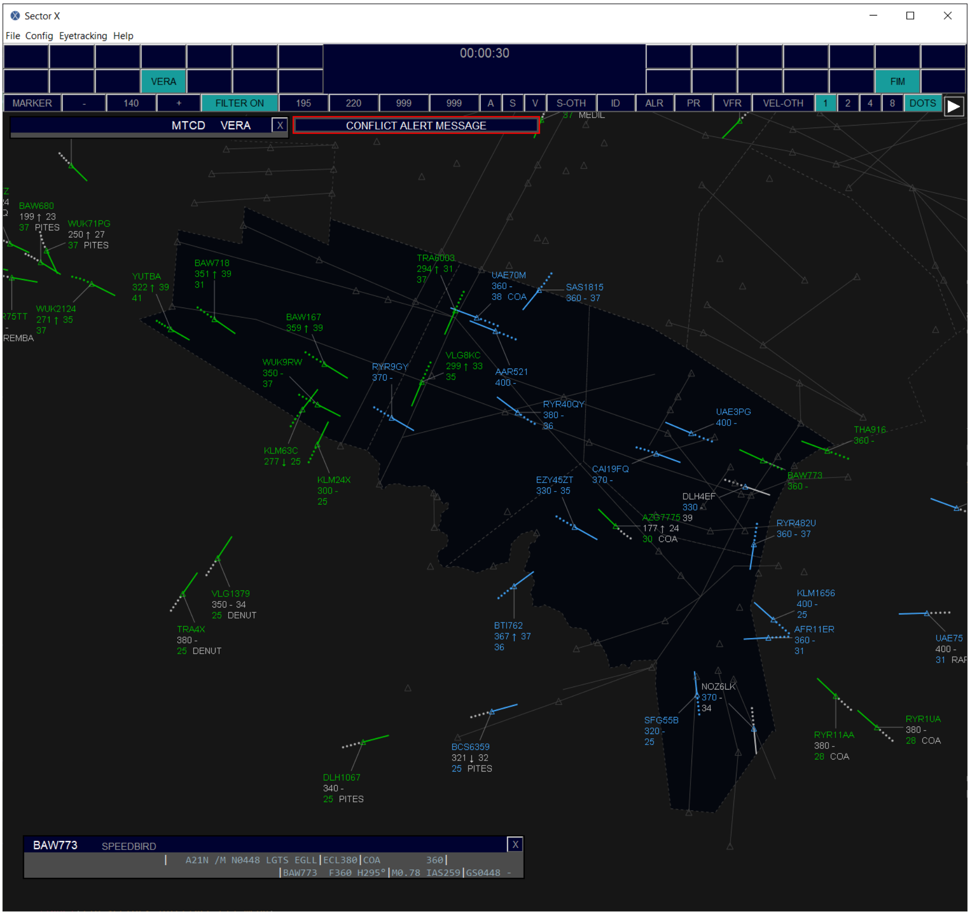


Figure A.1: SectorX in MUAC style.

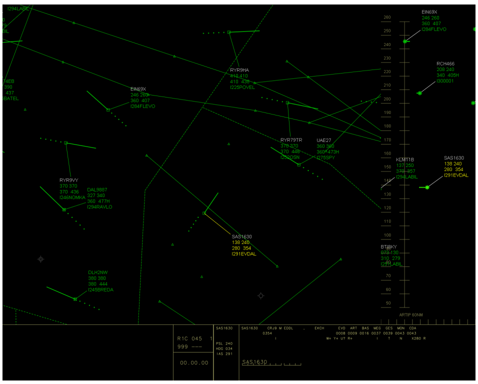


Figure A.2: SectorX in LVNL area control style.

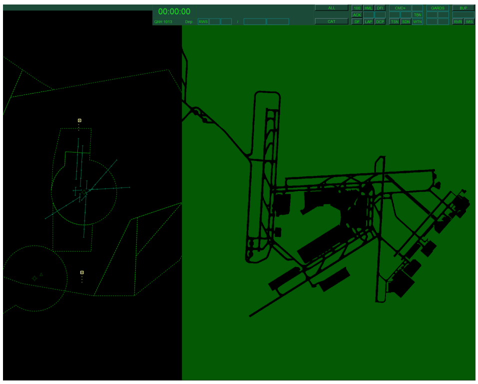


Figure A.3: SectorX in LVNL tower style.

A.3.1 Radar menu

The radar menu at the top of the screen provides a series of options to adjust the radar display. Only a subset of the buttons, shown in Figure A.4, was functional in this thesis' experiments. The speed vectors could be lengthened beyond the standard 1 minute to 2, 4 or 8 minutes, and the ATCOs could toggle the history dots that trail each RPS. While these dots can be used to judge ground speed and turning behavior, some ATCOs prefer to hide them to declutter the screen.

Figure A.4: Part of the MUAC radar menu showing the buttons that were active in the experiments.

A.3.2 Flight information management (FIM)

The flight information management (FIM) window shows flight plan information of a selected flight, together with Mode S transponder data (pilot selected flight level and heading, actual Mach number and indicated airspeed) and derived groundspeed and vertical speed (Figure A.5). In this thesis' experiments, the ATCOs mainly used this window to see a flight's aircraft type, destination, groundspeed and vertical speed.

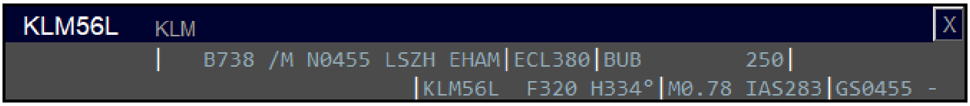


Figure A.5: MUAC flight information management (FIM) window for a KLM Boeing 737-800 flight from Zurich (LSZH) to Amsterdam (EHAM). The pilot has selected FL320 in the autopilot and the aircraft is flying on a heading of 334° with an indicated airspeed of 283 knots (Mach 0.78) and a groundspeed of 455 knots with no vertical speed. Its exit coordination point (XCOP) is BUB, where the flight needs to be at FL250.

A.3.3 Interactive labels and clearance menus

Each flight has an associated label on the plan view display, containing label items as illustrated in Table A.1. Depending on the state of the flight and its allocation to either human or automation, the label items and colors can change. Table A.2 shows examples of labels as encountered in the experiments of this thesis. All labels (and radar position symbols, RPS) in green relate to human-controlled flights, while blue items denote flights (suggested to be) allocated to the automation. When hovering the mouse over a label, a magnified label appears showing additional information. Only the label items needed for the experiments were included (e.g., due to the absence of speed control the speed label item was omitted). The standard, non-magnified, labels follow a minimal information approach, hiding any non-relevant information. This allows an ATCO to use them as a todo-list (i.e., labels with little items require little action).

Table A.1: MUAC flight label items, with actual flight level (AFL), cleared flight level (CFL), cleared heading/route point (HDG), transfer flight level (TFL), and exit coordination point (XCOP).

CALLSIGN
AFL - CFL HDG
TFL XCOP

Table A.2: MUAC flight labels in SectorX as used in the experiments of this thesis. White and gray colors have been replaced with black here for print clarity.

Situation	Standard label	Magnified label
Unconcerned flight. The flight is currently controlled by an adjacent sector and will not be transferred to the ATCO, but its flight plan lead/leads through the sector.	KLM123 340 -	KLM123 340 – 34 34 REDFA
Previous sector is controlling the flight, but it will be transferred to the ATCO's sector at some point. Color indicates the suggested allocation.	KLM123 340 – REDFA	KLM123 340 – 34 34 REDFA
	KLM123 340 – REDFA	KLM123 340 – 34 34 REDFA
Flight has been assumed by the ATCO, is flying level at its designated TFL (FL340) and on own navigation towards the XCOP (REDFA). No further action required.	KLM123 340 –	KLM123 340 – 34 34 REDFA
Flight is flying level at FL310, but needs to be at FL340 when exiting the sector. A climb instruction is required.	KLM123 310 – 34	KLM123 310 – 31 34 REDFA
Flight is climbing from FL314 to its cleared FL340, which is also its TFL. No further action required.	KLM123 314 ↑ 34	KLM123 314 ↑ 34 34 REDFA
Flight is at the correct flight level, but is flying on a heading (230°), meaning that it might not end up at the XCOP (REDFA).	KLM123 340 – H REDFA	KLM123 340 – 34 H230 34 REDFA
Flight has been transferred to the next sector.	KLM123 340 –	KLM123 340 – 34 34 REDFA
	KLM123 340 –	KLM123 340 – 34 34 REDFA
Flight is under automation control in the ATCO's sector. It has been cleared to descend to FL360, but needs to descend further to FL340 before leaving the sector.	KLM123 379 ↓ 36 34	KLM123 379 ↓ 36 34 REDFA

All label items are interactive, meaning that ATCOs can click on them with their mouse to open associated clearance menus. For example, clicking the CFL on the second label line will open a CFL menu, as shown in Figure A.6. The mouse is automatically centered on the TFL that the flight needs to be at when transferred to the next sector, although a different level can be selected by moving the mouse or using the scroll wheel to cycle through the list. Clicking any flight level automatically moves the mouse to the 'execute' button, such that the selected clearance can be executed swiftly with minimal mouse movement. Similar menus exist for heading and route clearances. In Chapter 3, the ATCOs could allocate flights to themselves or the automation by clicking the callsign item (see Figure 3.6).

The provided inputs are used to keep track of what clearances have been given, and allow subsequent sectors to see what an incoming flight has been cleared to. In addition, support systems use the inputs for their alerting systems, for example when a flight deviates from the CFL or route. In reality, the inputs provided through the clearance menus are primarily relayed to the aircraft via voice-based radio transmissions (R/T), with the exception of certain CPDLC instructions. Throughout this thesis, all clearances were transmitted via CPDLC to remove the need for pseudo-pilots and voice R/T.

Figure A.6: MUAC clearance menus in SectorX, from left to right: flight level (2x), heading and route.

A

A.3.4 **Verification and advice (VERA)**

MUAC's verification and advice (VERA) tool provides an easy-to-use measure of the predicted minimum horizontal separation between two flights. In other centers similar tools are known under the name 'horizontal scanning tool' (HST, Corver and Aneziris, 2015). Figure A.7 shows an example, where two flights have a minimum separation of 2.6 NM, which will be reached in 2 minutes from now. The predicted position of both flights at that moment is shown by the amber 'dagger' symbols. ATCOs use VERA to verify separation and/or to inspect conflict geometries to, e.g., decide which flight can be turned to fly 'behind' the other flight. VERA is shown on request when the ATCO 'connects' two flights with the mouse and can be toggled or locked to monitor a flight pair over a prolonged time. In the experiment of Chapter 6, the automation could autonomously enable VERA for a mixed flight pair, like ATCOs at MUAC can do to inform their colleagues of an imminent conflict.

Figure A.7: MUAC VERA.

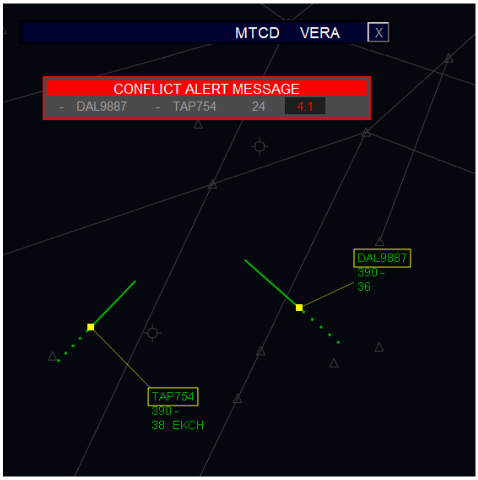


Figure A.8: MUAC STCA.

A.3.5 **Short-term conflict alert (STCA)**

Short-term conflict alerts (STCA) were implemented according to the example from Eurocontrol (2007) and were shown on the HMI as illustrated in Figure A.8. To alert the ATCO, both flights' callsigns are outlined in yellow, while their respective RPS continuously flash yellow and red. The 'conflict alert message' window provides further information on the nature of the conflict. In this example, the two flights are currently 24 NM apart and heading for a minimum separation of 4.1 NM.

STCA automatically triggers whenever two flights are predicted to have a horizontal separation of less than 5 NM horizontal and simultaneously less than 1,000 ft vertical separation. Unlike VERA, STCA considers the vertical dimension. It takes the cleared flight level into account as well, to prevent nuisance alerts when flights are expected to level off before bursting through an occupied flight level.

A.3.6 Label decluttering

For the experiment in Chapter 6, a label decluttering algorithm devised by Reek (2010) was implemented because the relatively large number of flights would lead to many overlapping labels. Manually moving the labels was expected to add an excessive amount of workload and would likely confound eye tracking and control activity measures. Furthermore, the ATCOs participating in the experiment of Chapter 3 unanimously indicated that labels should (mostly) be decluttered automatically (Figure 3.10), like already done on the actual MUAC HMI.

The chosen implementation makes use of nine possible label locations around the RPS, which are evaluated for each flight on each radar update. A cost (C) is calculated for each position according to Equation (A.1) and the label is then moved to the position with the lowest cost. To prevent labels from jumping around on every update, a heavily weighted jitter penalty of 4 is added on each label movement, and subsequently reduced by 1 on each update (i.e., after four updates the penalty from a single movement has been removed). The overlap costs were as listed in Table A.3 and C_{distance} was simply the distance (in pixels) from the old to the new location of the label.

$$
\begin{aligned}
C &= 1000 \cdot C_{\text{jitter}} + 100 \cdot C_{\text{overlap}} + 10 \cdot C_{\text{angle}} + 2 \cdot C_{\text{distance}} \qquad \text{(A.1)} \\
C_{\text{angle}} &= ||\text{mod}(\text{track} - \text{leaderLineAngle} + 540, 360) - 180| - 135|
\end{aligned}
$$

Table A.3: Label decluttering overlap costs (C_{overlap}).

	Overlap	
	Label	Leader line
Symbol	40	Not implemented
Label	30	20
Leader line	20	25
Speed vector	15	5

If the ATCOs preferred a different position, they could click the aircraft symbol to rotate the label to the next position. Following feedback from the participating ATCOs, this has since been changed to immediately jump to the position with the next lowest cost, rather than cycling through all available positions which may or may not have an even higher cost. In the experiment, ATCOs occasionally had to click multiple times until the label reached a low-cost position where it did not overlap with other labels.

A.4 Aircraft performance models

SectorX supports three types of aircraft performance models, ranging from very basic to an extensive parametric model widely used in industry. The latter was used in all experiments of this thesis, but more basic models can be sufficient for evaluations with ATC novices (van Paassen et al., 2023).

A

Basic

The basic model allows users to specify their own performance models, based on XML files with parameters for acceleration, deceleration, vertical rates and bank limits per aircraft type. These values are all independent of altitude and airspeed and therefore mostly useful for simulations that are limited to the horizontal plane.

Generic

A slightly more advanced generic model includes three aircraft types: light, medium and heavy. Per type, average climb/descent rates, and minimum and maximum speeds have been selected based on the BADA model for the Cessna 550, Boeing 737 and Boeing 747 respectively (Eurocontrol, 2012). Unlike the basic model, vertical rates do vary with altitude to some extent. Bank angle is limited to 30 degrees for all types in all conditions.

BADA

The most advanced model is the Base of Aircraft Data (BADA) 3.10 parametric performance model (Eurocontrol, 2012), which provides lookup tables with reference values for circa 400 aircraft types. Apart from a performance model, BADA also supplies matching airline procedure models (e.g., climb and cruise speeds) and aircraft characteristics (e.g., wing area, turbulence category). Note that the airline procedure models are not in fact airline-dependent, but solely aircraft type dependent and merely approximate average airline practice.

The ATCOs participating in this thesis' experiments rated aircraft performance moderately realistic (Figures 3.19 and 6.25). Especially climb and descend speeds need further tuning as these were often too high compared to real life. Using standard climb/descend speeds instead of the BADA provided maxima is expected to provide a significant improvement.

A.5 Automated ATC agent

SectorX supports the implementation of automatic solvers that can act as a digital 'ATCO'. The solver used in the experiments of Chapters 3 and 6 would follow three consecutive algorithmic steps, for which pseudo-code is presented below. The steps were repeated every 30 seconds for all flights. For the fast-time simulation that was used in designing the interaction-based allocation scheme of Chapter 5, the algorithms were modified to ignore any conflicts.

Algorithm 1 solve conflicts (only for fully automated pairs),

Algorithm 2 send direct-to as far as the XCOP if possible (not used in Chapter 3),

Algorithm 3 clear to next available flight level, closest to transfer flight level.

When checking for conflicts and the safety of clearances, the same routines as STCA (see Appendix A.3.5) were used, but with 8 minutes look-ahead time and an additional separation buffer of 2 NM. Conflicts were solved pairwise, meaning that the solver had no notion of multi-flight conflicts and would not necessarily pick the best solution in such a case. The pseudo-code shown here reflects the final experiment in Chapter 6. The preliminary experiment from Chapter 3 used a slightly simplified version.

A

Algorithm 1 Solve conflicts

if one or both flight(s) is/are climbing or descending **then**
    Stop climb/descend of one of the flights
else if both flights are flying level **then**
    if descend available **then**
        Descend either flight 1000 ft
    else
        Climb either flight 1000 ft
    end if
end if

The automation would always try to send a flight as direct to its XCOP as possible. However, flights were not allowed to clip the sector boundaries, meaning that only trajectories fully within the controlled sector boundaries were accepted. The process was repeated for consecutive route points in the flight plan, starting from the XCOP and working towards the current position, until a conflict-free and non-clipping route was found. Note that the algorithm only checks whether the trajectory was conflict-free for the duration of its look-ahead time.

Algorithm 2 Clear direct-to if possible

furthest route point = XCOP
for furthest route point **do**
    if conflict free **and not** clipping sector border **then**
        Clear flight direct-to
        return
    end if
    furthest route point = next route point closer to current position
end for

A

Automation would only clear flights to a flight level that could be reached within its look-ahead time. This prevented the blockage of excessive altitude bands by giving flights an unrealistically large climb/descend instruction. The automation would issue the next clearance (if needed) when a flight was within 3,000 ft of the previously cleared level, to prevent continuous clearances as the look-ahead window progressed in time.

Algorithm 3 Finding best available flight level

ROCD $= \frac{|\Delta FL|}{d_{todestination}}$
if ROCD < -1900 ft/min **then**
    Descend aircraft
end if
if not safe **then**
    Check if altitude is safe 1000 ft lower
end if

B

Experiment briefing and questionnaire

As an example, this appendix contains the experiment briefing and questionnaires for the experiment from Chapter 6. The briefings and questionnaires for the other experiments were comparable.

B.1 Briefing

Experiment briefing

Empirical Analysis of Flight Allocations in a Shared Airspace

Responsible researchers: Gijs de Rooij (G.deRooij@tudelft.nl) and Clark Borst (C.Borst@tudelft.nl)

Introduction

Dear participant,

Thank you very much for taking part in this research for my PhD project! With this simulation experiment, we would like to gain insights into a future concept of operations, where the 'basic' flights in your airspace may be delegated to a computer, so you can focus on the more interesting and complex flights.

The experiment will take at most 1.5 hours. This document contains important information to prepare you. Please read it advance so you know what to expect.

Do not hesitate to contact me in case of any doubts or questions!

Kind regards and see you soon,

Gijs de Rooij
PhD Candidate in Aerospace Engineering

Experiment set-up and procedure

Simulator
The simulator that we use in this experiment is developed by TU Delft, and designed to resemble the MUAC interface. For the sake of the experiment there will be several differences that we will let you experience at the start of the experiment. Most importantly, you only need to input clearances in the system (simulating CPDLC) as there is no voice R/T.

Scenarios
You will be presented with two scenarios of ca. 25 minutes each, based on a recent radar sample. You are responsible for the entire Brussels sector group. However, roughly half of the flights in this airspace will be delegated to the computer, meaning that you have no control over them. These 'automated flights' are coloured blue, as shown in Figure 1 (your own flights are green). In either scenario, this distribution is based on different rules, which will be introduced beforehand.

Your task
It is your task to issue any clearances that you deem necessary to ensure that all flights can safely reach their XCOP at the correct TFL. You can give altitude, heading or route/direct clearances. Speed is at pilot discretion and cannot be modified by you. You can preview the planned trajectories and use VERA to assess potential conflicts. Blue flights will automatically be assumed, controlled and transferred by the computer.

Automation
The computer will climb blue flights as early as possible, descend them as late as possible and send them on direct routes to the XCOP when able. It has a lookahead time of 8 minutes and will take care of automatically preventing and/or solving conflicts between blue flights in this time frame. It will not steer blue flights into conflict with green flights, but if such conflicts do occur (after 8 minutes, or because you modified a green flight), the computer will **not** solve them. You are responsible for solving these 'mixed' conflicts, by issuing a clearance to the green flight. Automation will notify you when such conflicts are predicted, 8 minutes before the LOS, by marking the conflicting pair with VERA.

Fig. 1: Simulator interface

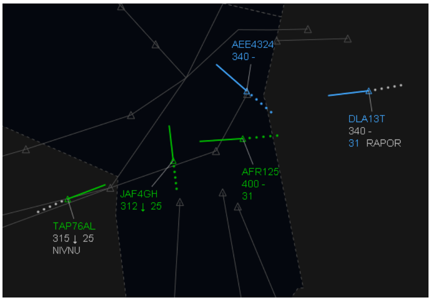


Fig. 2: Eye tracker

Your rights

Participating in the experiment is on voluntary basis. You may cancel your participation at any moment for any reason, even during the experiment. This has no consequences for you. Your personal performance is not part of the study, there is no wrong or right, and there is no competition between you and your colleagues.

Data that are collected during the experiment will be anonymised before they are stored by assigning a random identifier to you. Only the principal researcher (Gijs de Rooij) and his daily supervisor (Clark Borst) can link the data to you personally. This relation will never be made available to others (neither to MUAC). By participating, you give consent to publishing the data in anonymised form.

In the experiment, you will be asked to wear an eye tracker in the form of a pair of glasses (Figure 2). This eye tracker records videos of both your eyes. After the experiment, we can derive from these videos where you were looking on the screen. The raw videos will be deleted as soon as this info is derived, but in any case no later than 3 months after the experiment date.

At the start of the experiment you will be asked to sign a consent form to make sure that you have read and understood what participating in the experiment means.

Important information if you wear eyeglasses and/or make-up

For the best results, it is advised that you **do not wear your own glasses in the experiment**, as they may interfere with the recordings. **Contact lenses are fine**. In addition, **your eyes and area around them is preferably free of make-up** as much as possible, as this can cause reflections and therefore noisy data.

B.2 Questionnaire

The questionnaires were administered through an online Qualtrics environment before, during and/or directly after the experiment.

Introduction

This questionnaire consists of four pages, with questions about:

1. You and your professional experience
2. The flight allocation algorithms
3. The automation
4. Simulator and experiment fidelity

Every page ends with a text box where you can type any additional comments you may have that did not fit the preceding questions. You can use the buttons at the bottom of the screen to go back and forth.

You and your professional experience

Age	☐
Years of professional ATCO experience	☐

Have you been involved in developing and/or testing an ATC environment where part of the flights are controlled by automation? Can you very briefly explain what kind of research/development that was?

☐

Flight allocation algorithms

In the experiment you experienced two scenarios with two different flight allocation algorithms. The following questions refer to this as the first and second algorithm respectively.

	Dislike a great deal	Dislike somewhat	Neutral	Like somewhat	Like a great deal
First algorithm	○	○	○	○	○
Second algorithm	○	○	○	○	○

B

Please drag the slider to indicate to which algorithm the following statements applied the most. Place the slider in the middle (neutral) if it was the same in both scenarios.

	First algorithm	Neutral	Second algorithm
Lower workload			
Consistent workload			
Better situation awareness			
Automation and I were a team			
I could organize my work the way I wanted			
Automation was helpful			
I could solve conflicts efficiently			
Automation and I kept the sector safe			
Automation and I managed the traffic efficiently			

What did you think were the 'conditions' under which the algorithm in the *first* scenario would allocate a flight to **automation**?

Were there any things that you particularly liked or disliked about the flight allocation in the *first* scenario?

What did you think were the 'conditions' under which the algorithm in the *second* scenario would allocate a flight to **automation**?

Were there any things that you particularly liked or disliked about the flight allocation in the *second* scenario?

If you were to design an automatic algorithm for allocating flights to either automation or a human ATCO, what would that allocation look like? What should the algorithm take into consideration?

Is there anything else that you would like to share about the allocation algorithms?

Automation

The questions on this page relate to the functioning of the automation itself, which was independent of the flight allocation algorithm.

How much trust did you have in the automation?

None at all	A little	A moderate amount	A lot	A great deal
○	○	○	○	○

Please indicate to what extent you agree or disagree with the following statements about the automation.

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
It was reliable	○	○	○	○	○
It was predictable	○	○	○	○	○
It was consistent	○	○	○	○	○

Is there anything else that you would like to share about the automation?

Simulator fidelity

How would you rate the realism of the following simulation aspects? Please compare the simulation to the actual MUAC working position and live traffic. Try to exclude details from your judgment that might be missing but were not relevant for the current experiment.

	Unrealistic	-	Acceptable	-	Perfect
Interface (look and feel)	○	○	○	○	○
Label decluttering	○	○	○	○	○
Aircraft behavior (vertical rates, turns)	○	○	○	○	○
Traffic scenario (density, routes)	○	○	○	○	○

Were there any features that you missed while performing the experiment?

B

Do you have any other comments about the simulation? What did you really like, or what can we improve for future experiments?

Anything else?

Is there anything else that you would like to share with us? Something that was not treated in any of the questions? Or would you like to share your own experiences or ideas?

Please make sure you are ready to submit your answers. Pressing the 'Next page' button will close the survey.

C

Preventing scenario recognition

In academic air traffic control research, traffic scenarios are often repeated to increase the sample size and enable paired-sample comparisons, e.g., between different display variants. This comes with the risk that participants recognize scenarios and consequently recall the desired response. This chapter provides an overview of mitigation techniques found in literature and concludes that rotating scenario geometries is most frequently used. The potential impact of these transformations on participant behavior, as described in this chapter, is however not sufficiently addressed in most studies. As an example, eye tracking data is analyzed from the experiment of Chapter 4 in which ten professional air traffic control officers were each presented with three repetitions in various rotations of several distinct scenarios. Results imply that researchers wishing to repeat scenarios should more carefully consider whether mitigation techniques might have an impact on their results.

Parts of this chapter have been published in:

de Rooij, G., Borst, C., van Paassen, M.M., and Mulder, M. Preventing Scenario Recognition in Human-in-the-loop Air Traffic Control Research. In *22nd International Symposium on Aviation Psychology*. Rochester, NY, USA, 2023a

C.1 Introduction

In air traffic control officer (ATCO) training and airspace redesign trials, simulation scenarios are designed to be as realistic as possible, with many different flights over a prolonged period of time. High face validity enables the ATCOs to execute their tasks as they would in an operational setting. Academic research, however, often benefits from simplified, more constrained scenarios that are presented to novices or experts while tracking their behavior e.g., when using different display variants. Constructing alike scenarios, where the scenario itself is not an independent variable, is a major task, requiring considerable effort and input from subject matter experts. As an alternative, identical traffic scenarios are, therefore, often repeated to obtain paired-samples at the risk of scenario recognition. Depending on the aim of the study, this can be undesirable as participants may recall their earlier responses rather than coming up with an independent solution, aggravating learning effects. This applies especially to studies that measure ATCO consistency, such as in the personalization of conflict resolution advisories (Westin et al., 2016a). Finding a balance between using alike scenarios and preventing recognition is not trivial.

This chapter, for the first time, provides an overview of techniques used to mitigate scenario recognition in existing air traffic control (ATC) studies. A straightforward and frequently employed method is to rotate and/or mirror scenarios. While these transformations result in identical scenarios in terms of conflict angles, traffic densities and patterns etc., the change in orientation may unconsciously impact participant behavior. This may not reveal itself in the final outcome, e.g., solving a conflict, but it can elicit different visual scan patterns to arrive at this outcome. Visual search is an essential process that ATCOs use to continuously update their mental picture (Fraga et al., 2021). Changes in this process may lead to faster or slower conflict detection in otherwise identical scenarios, affecting related objective measures. Furthermore, perceived workload may be affected (e.g., due to unusual traffic directions, especially for experts) and action sequences or conflict resolutions might change due to different fixation orders.

These effects are, to the best of our knowledge, not sufficiently identified and recognized in literature. Authors often merely mention that scenarios are transformed to 'prevent recognition' without further detailing their considerations or the transformation's implications. In addition to our literature survey on mitigation techniques, we therefore analyze eye tracking data from a previously executed experiment that featured scenario transformations (Chapter 4). The data consists of ten professional ATCOs who each performed conflict detection and resolution in 15 distinct scenarios, of which five were selected for this analysis. Each scenario was presented three times to them with different transformations. By comparing the order in and speed at which flights were fixated, we empirically describe the participants' behavioral consistency when presented with transformed repetitive scenarios. To conclude we argue on the implications that researchers should consider when repeating scenarios, based on these initial findings.

C.2 Mitigation techniques

A literature survey resulted in the identification of three categories of techniques to prevent scenario recognition, explicitly described in 20 ATC studies and summarized in Table C.1: geometric, textual and temporal. Most studies used a combination of techniques, with rotating scenarios as the most popular technique, employed in 15 studies.

Table C.1: Scenario recognition mitigation techniques explicitly mentioned in existing research.

	Geometric		Textual		Temporal	
Study	Rotation	Mirroring	Renaming callsigns	Renaming waypoints	Time shifting	Reordering
Abdul Rahman (2014)	✓	-	-	-	-	-
De Albuquerque Filho et al. (2008)	-	**?**	-	-	✓	✓
Borst et al. (2017)	✓	-	-	-	-	✓
Borst et al. (2019)	✓	✓	-	-	-	-
Cummings et al. (2005)	✓	-	-	-	-	-
Harrison et al. (2014)	-	-	✓	-	-	-
Hilburn et al. (2014)	✓	-	-	✓	-	-
IJtsma et al. (2022)	✓	-	-	-	-	-
Jans et al. (2019)	✓	-	-	-	-	-
Jasek et al. (1995)	-	-	-	-	-	✓
Jha et al. (2011)	✓	-	✓	✓	-	-
Kim et al. (2022)	✓	-	✓	✓	-	-
Klomp et al. (2016)	✓	-	✓	-	-	-
Major and Hansman (2004)	✓	✓	-	-	-	-
Metzger and Parasuraman (2006)	✓	-	-	-	-	-
Rovira and Parasuraman (2010)	✓	-	✓	✓	-	-
Sollenberger and Hale (2011)	-	-	✓	-	-	-
Ten Brink et al. (2019)	✓	-	-	-	-	-
Trapsilawati et al. (2021)	✓	-	✓	✓	✓	-
Wilson and Fleming (2002)	-	-	✓	-	-	-
Number of studies	15	2/3	8	5	2	3

Geometric When a scenario is rotated or mirrored, its (objective) taskload formed by the traffic density, conflict geometries etc. remains the same, but its (subjective) workload might change. Especially with experts, accustomed to traffic streams from certain directions, changing the principal axis can have an impact on their perceived workload, as it requires a change in scan pattern.

Geometric transformations can only be done when the sectors are relatively symmetric, which is generally not the case in operational environments. Furthermore, on a widescreen monitor, rotations other than 180° may result in a reduced look-ahead range for flights coming towards the sector. Square-shaped monitors (or simulated windows), as found in many ATC centers, eliminate this problem. Only rotation multiples of 90° were found in the studies, presumably because this generates sufficient transformations and is easy to execute. De Albuquerque Filho et al. (2008) mention that they 'invert the route structure', without further detailing what is meant by that.

Textual Changing callsigns and waypoint names is a simple technique that can be widely applied, does not change the taskload and has proven to be sufficient on its own in some cases, such as the study by Wilson and Fleming (2002). When realistic callsigns and aircraft performance data are used, the callsign should match the flight's characteristics (e.g., no big airliner for small airlines or non-standard destinations). Similarly, when using operational airspaces, waypoints may need to be left unaltered to match operational routes. Neither are a problem when using airspace-naive novices.

Temporal Shifting occurrences of, for example, conflicts in time is a feasible technique for relatively long scenarios, where chunks of traffic entering the sector can be shuffled (De Albuquerque Filho et al., 2008, Trapsilawati et al., 2021). Such temporal transformations do, however, risk ignoring cognitive built-up and its associated impact on (perceived) workload. This technique is, therefore, mostly used to construct realistic scenarios from recorded flight data, by shifting flights to create a plausible scenario that is denser or has more conflicts than the recording.

When an experiment consists of multiple scenarios per test condition, their order can be changed. If, for example, display variants are tested that are sufficiently distinct from each other, participants may be predominantly occupied by the changed visuals and/or tasks, making it even less likely for them to recognize repeated scenarios at all (Jasek et al., 1995).

An extreme case of re-ordering chunks of traffic is to add dummy scenarios in between measurement scenarios, as done by Borst et al. (2017). If planning allows, measurements for each participant can even be split over multiple days. This requires good planning (difficult when using experts) and is more prone to introducing confounds due to a lack of control over variables such as participant energy levels or between-session (professional) experiences. It is therefore not often used, except in longitudinal studies such as by Hilburn et al. (2014).

A technique not explicitly found in literature is the shifting of all flights up or down in altitude. The individual contribution might be marginal, as humans predominantly recognize plan-view patterns, but in combination with other techniques it can require participants to not completely rely on their memory. Care must be taken not to alter the altitudes too much, as changes in flight level have an effect on ground speeds and thus closing rates, impacting the time a loss of separation occurs and/or conflict warnings will be issued.

C.3 Data description

As an example of the potential impact of scenario transformations, we revisit and analyze eye tracking data from a previously executed experiment designed for task analyses (Chapter 4). To prevent scenario recognition, it involved static scenarios featuring several geometric and textual transformations, dummy scenarios and a varying scenario order.

C

C.3.1 Participants and apparatus

Ten professional en-route ATCOs (age: M = 43.6, SD = 7.1, years of experience: M = 20.0, SD = 6.5), from Eurocontrol's Maastricht Upper Area Control Centre (MUAC) voluntarily participated in a simulator experiment, as approved by the Human Research Ethics Committee of TU Delft under number 2754. All participants provided written informed con-

Figure C.1: Experiment set-up with participant (left) and observer (right) positions.

sent. A TU Delft-built medium-fidelity simulator was designed to mimic the MUAC interface on a 1920 x 1920 pixels 27″ display with a computer mouse for control inputs, shown in Figure C.1. Although the scenarios were static, participants could measure predicted minimum separation between flights and display extended flight labels.

Gaze data was recorded using a head-worn Pupil Labs Core eye tracker (Kassner et al., 2014) with Pupil Capture v3.5.1. The forward-facing scene camera recorded at 30 Hz and the pupils were recorded at 120 Hz. Eight AprilTag markers were placed along the edges of the screen to relate gaze to screen pixels. Clusters of gaze points that were close in location and time were classified as fixations through the Python version of I2MC by Hessels et al. (2017), with a minimal duration threshold of 60 ms as used by Fraga et al. (2021). The fixations were correlated to flights by drawing voronoi-like areas of interest around each flight's symbol, speed vector and label.

C.3.2 **Scenarios**

The ATCOs assessed 15 distinct static scenarios that were shown three times, each time with different transformations. They featured an artificial, octagonal 80 x 80 NM sector, with four waypoints in the cardinal directions. This made sure that the ATCOs would not fully rely on their trained scan patterns and that repetitions could not be recognized based on the sector shape. Two, three or four flights were present on direct routes to their exit points. Variants were created by applying any (combination) of the following:

- Rotation: 90, 180 or 270 degrees,
- Mirroring: flipping along the x- or y-axis,
- Altitude shift: all flights up or down by 1,000 or 2,000 ft.

Callsigns were randomized for all variants and flight labels were always placed at a 90 degree offset to the direction of travel. Figure C.2 shows an example of a scenario with corresponding transformations. Note that flights in the center of the sector were invariant to all geometric transformations and always appeared at the same location on the screen while their label was moved to match the new direction of travel. All participants got to see the same order of transformations, but the scenario ordering was counterbalanced between them to account for learning effects. This ordering of scenarios was defined in the previously executed experiment and might, in hindsight, have been suboptimal for the current study. The experiment started with six training scenarios.

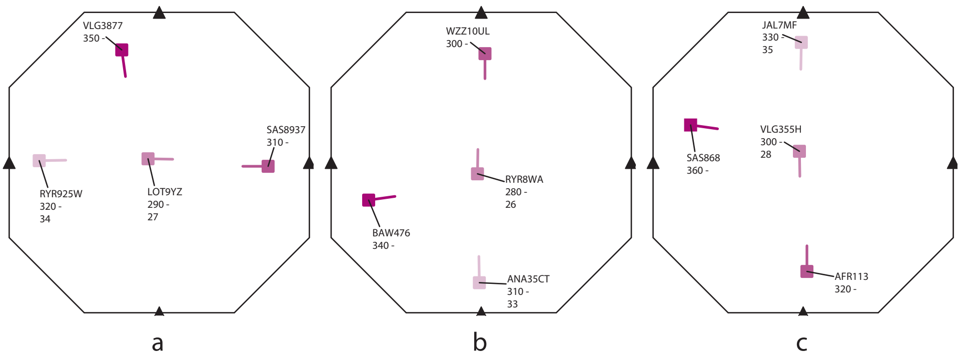


Figure C.2: Three transformations of Scenario 5. Colors relate to the same flights in each transformation.

C.3.3 **Participant task**

Participants were asked to first indicate for each scenario whether there were any conflicts and to consequently solve these through altitude clearances only. Some flights had to leave the sector at a different flight level, requiring a clearance that would generally also solve any conflict(s). An intermediate level was needed in some cases to not create a conflict. If there were no (remaining) conflicts and all flights were at or cleared to the correct flight level, the ATCO could advance to the next scenario by clicking a button in the lower right corner of the screen. This button was carefully placed to ensure a common first fixation point, not related to any flights, when a scenario loaded. As the experiment was designed for Chapter 4, it involved a second phase that was not included in the present analysis.

C.4 **Results and discussion**

After the experiment, some ATCOs mentioned that they did recognize the repetition of certain conflict geometries, but none of them recalled that they were identical scenarios apart from the applied transformation(s). Our present analysis stays away from concluding whether the recognition mitigation has worked and instead focuses on the consistency of fixation behavior. For the fixation measures, only the five scenarios containing four flights are included, because for the other scenarios with two or three flights these measures would be less robust. Since the ATCOs showed vastly individualized behavior, no between-participant comparisons are performed and all observations discussed here relate to the three scenario repetitions per individual.

C.4.1 **Conflict assessment**

The participating ATCOs were not always consistent in what situations they flagged as conflict. Figure C.3 shows the ATCOs conflict assessment for each of the 15 scenarios. Nineteen (13%) out of 150 scenarios (15 per ATCO) showed ambiguous results, with ATCOs changing their opinion when presented with a transformed version of an otherwise identical scenario. In 11 cases the first repetition was the odd one out, in five cases the middle and in three the last. ATCO 3 consistently flagged all flights that would cross another flight's path, if they were to be cleared to their transfer flight level, despite being briefed to only consider the current and cleared flight levels.

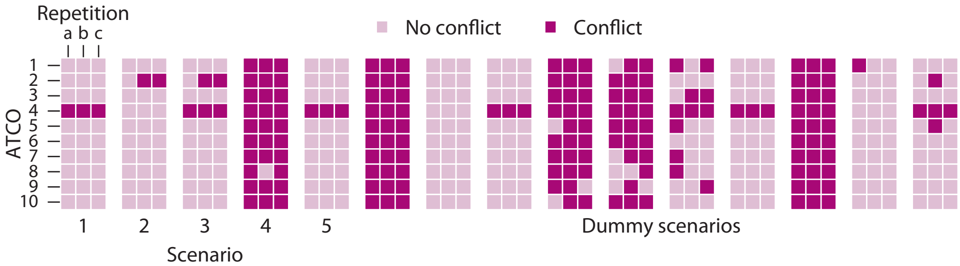


Figure C.3: Conflict assessments, including the 10 dummy scenarios.

C.4.2 Fixation order

Conflict detection time is directly driven by the order in which flights receive attention, especially when scenarios include many flights. After all, if an ATCO fixates flights in a different order, he/she might observe a conflicting pair earlier or later. To this end, Figure C.4 shows for each scenario's three repetitions the flight that was first fixated by each ATCO. The level of consistency, in terms of identical first fixations for all three transformations (visible as a row of three similarly colored squares), varied per ATCO from zero (ATCOs 5 and 10) to three scenarios (ATCOs 7 and 8). A similar variance can be seen between the scenarios, with consistent first fixations for one (Scenarios 1 and 2) to five (Scenario 5) ATCOs. This suggests that the rotations may have had an impact on the fixation order, and that this can differ per individual and traffic layout. On closer inspection, in 80% of the runs, the first fixated flight in Scenario 5 was in the center of the sector (and therefore in the exact same location for all repetitions). Conversely, Scenario 1, the only one with no flight near the center, shows the lowest level of consistency.

To illustrate individual differences, complete orders of fixation for two ATCOs on either extremes of the aforementioned consistency scale are shown in Figure C.5. Note how ATCO 8's complete fixation sequence is consistent for all variants of Scenario 3. This, in combination with the inconsistent fixation orders seen in other scenarios or with other ATCOs, further hints at a non-negligible influence of scenario rotation on the processing of traffic scenarios. For more insight into the relevant mechanisms, an analysis of scan patterns at different transformations would be useful, but this requires scenarios with more flights. The static, low density scenarios used in this study imply that the results are not necessarily applicable to dynamic and/or denser scenarios.

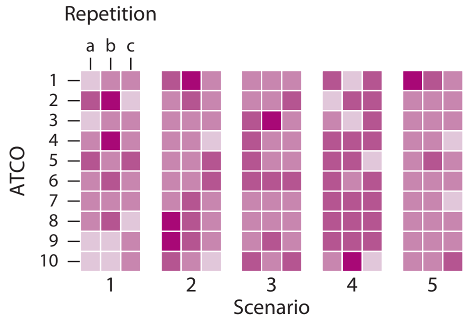


Figure C.4: First fixated flight per ATCO. Colors represent specific flights in a scenario (see Figure C.2 for Scenario 5).

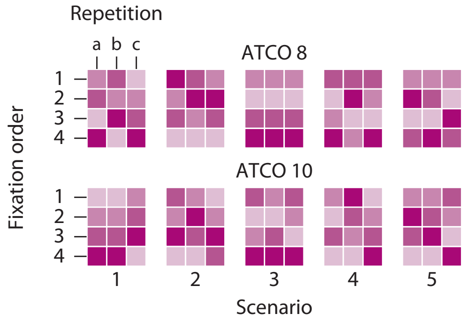


Figure C.5: Complete flight fixation orders of two ATCOs. Colors represent specific flights in a scenario (see Figure C.2 for Scenario 5).

C.4.3 Fixation speed

To further illustrate the potential influence of rotations on fixation sequences and duration, Figure C.6 shows the standardized time till specific flights in Scenarios 3 and 5 had been first fixated. Results imply that the rotational-influence on this measure is dependent on the researcher's flight of interest. This is most visible in Scenario 5b, where Flight 1 shows significantly different means compared to the other two rotations. Akin to the fixation order, differences between individuals are again considerable, reflected in the wide spread of most data.

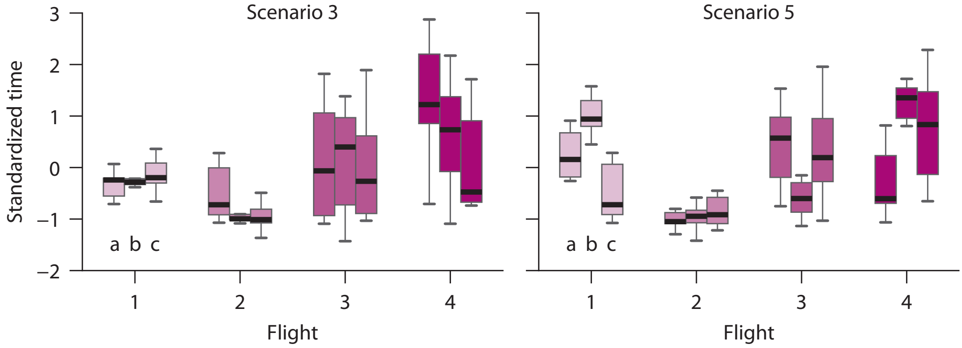


Figure C.6: Standardized (per ATCO) time till flights have been first fixated in two scenarios, split per transformation. Colors represent specific flights in each scenario (see Figure C.2 for Scenario 5).

While the order of scenarios was counterbalanced between the participants, the order of their repetitions was not (i.e., all ATCOs first saw a, followed by b and then c). While this resulted in a clearly visible reduction in total fixation time over the three repetitions, this reduction is not (always) reflected in the results presented here. We therefore conclude that this speed-up was mostly caused by the ATCOs getting more acquainted with the task at hand and advancing to the next scenario, rather than recognizing the specific scenarios. To further isolate the effect of purely the transformation, future studies should include duplicate scenarios where no transformation has been applied.

C.5 Conclusions

Scenario transformations such as rotation and mirroring are proven techniques to create paired-samples in human-in-the-loop ATC research, but the potential impact on results is not always sufficiently recognized. We showed that the most popular technique, rotating scenarios, does risk eliciting different eye fixation behavior from participants, potentially confounding objective measures such as conflict detection time. Whether this is problematic strongly depends on the research question(s) at hand and requires careful consideration. No definitive conclusions regarding the size of these effects can be made on the basis of the limited analysis presented here. The first indications do warrant further research with more elaborate, potentially dynamic, traffic scenarios and a tailored experiment design.

C

References

Abdul Rahman, S.M.B. *Solution Space-based Approach to Assess Sector Complexity in Air Traffic Control*. PhD thesis, Delft University of Technology, 2014. doi:10.4233/uuid:abf9f743-ea5c-4995-9371-c0018150b0cd.

Ahlstrom, V. Human Factors Design Standard (DOT/FAA/HF-STD-001B). 2016.

Alaeddini, A., Erzberger, H., and Dunbar, W. Distributed Logic-Based Conflict Resolution of Multiple Aircraft in Planar En-Route Flight. In *AIAA Guidance, Navigation, and Control Conference*. American Institute of Aeronautics and Astronautics, 2011. ISBN 978-1-60086-952-5. doi:10.2514/6.2011-6659.

de Albuquerque Filho, E.A.F., Trabasso, L.G., Sandes, A., de Araújo, M., Li, L., and Hansman, R.J. Experimental Setup for Air Traffic Control Cognitive Complexity Analysis. In *Symposium of Operational Applications in Areas of Defense*, pp. 342–347. São José dos Campos, Brazil, 2008.

Alharasees, O., Jazzar, A., Kale, U., and Rohacs, D. Aviation communication: the effect of critical factors on the rate of misunderstandings. *Aircraft Engineering and Aerospace Technology*, 95(3):pp. 379–388, 2022. doi:10.1108/AEAT-02-2022-0052.

Andraši, P., Radišić, T., Novak, D., and Juričić, B. Subjective Air Traffic Complexity Estimation Using Artificial Neural Networks. *Promet*, 31(4):pp. 377–386, 2019. doi:10.7307/ptt.v31i4.3018.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.P., Granger, G., Benhacene, R., and Babiloni, F. Adaptive automation triggered by EEG-based mental workload index: A passive brain-computer interface application in realistic air traffic control environment. *Frontiers in Human Neuroscience*, 10, 2016. doi:10.3389/fnhum.2016.00539.

Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., and Scott, L.A. Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. *Human Factors*, 48(4):pp. 693–709, 2006. doi:10.1518/001872006779166280.

Bainbridge, L. Ironies of Automation. *Automatica*, 19(6):pp. 775–779, 1983. doi:10.1016/0005-1098(83)90046-8.

Balfe, N., Sharples, S., and Wilson, J.R. Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System. *Human Factors*, 60(4):pp. 477–495, 2018. doi:10.1177/0018720818761256.

Batra, A., Rajkumar, S., Nielsen, D.L., Clarke, S.S., Kalyanam, K., Tejasen, K., Ohsfeldt, M., and Copp, M. Document Classification Techniques for Aviation Letters of Agreement. In *AIAA Aviation Forum and Ascend*. American Institute of Aeronautics and Astronautics, 2024. doi:10.2514/6.2024-4358.

Baumgartner, M. The Organisation and Operation of European AIrspace. In A. Cook, ed., *European Air Traffic Management: Principles, Practice and Research*, chapter 1. Ashgate Publishing, 2007. ISBN 9780754672951.

Beadle, A. A Statement on the Future of Global Air Traffic Management by IFATCA. 2007.

Bekier, M., Molesworth, B.R.C., and Williamson, A. Defining the drivers for accepting decision making automation in air traffic management. *Ergonomics*, 54(4):pp. 347–356, 2011. doi:10.1080/00140139.2011.558635.

Bekier, M., Molesworth, B.R.C., and Williamson, A. Tipping Point: The Narrow Path between Automation Acceptance and Rejection in Air Traffic Management. *Safety Science*, 50(2):pp. 259–265, 2012. doi:10.1016/j.ssci.2011.08.059.

Billings, C.E. Human-Centered Aviation Automation: Principles and Guidelines. Technical Report TM-110381, NASA, 1996.

Billings, C.E. *Aviation Automation: The Search for a Human-Centered Approach*. CRC Press, Boca Raton, FL, USA, 1997. ISBN 9780805821277.

Bin Jumad, A.S., Tominaga, K., Yi, C.X., Duong, V.N., Itoh, E., and Schultz, M. Flow-Centric Air Traffic Control: Human in the Loop Simulation Experiment. In *IEEE/AIAA 42nd Digital Avionics Systems Conference*. 2023. doi:10.1109/DASC58513.2023.10311168.

Birkmeier, B., Korn, B., and Flemisch, F.O. First findings on the controller's mental model in sectorless air traffic management. In *IEEE/AIAA 30th Digital Avionics Systems Conference*. IEEE, Seattle, WA, USA, 2011. ISBN 9781612847979. doi:10.1109/DASC.2011.6096100.

Birkmeier, B., Tittel, S., and Korn, B. Controller Team Possibilities for Sectorless Air Traffic Management. In *Integrated Communications, Navigation and Surveillance Conference*. IEEE, Herndon, VA, USA, 2016. ISBN 9781509021499. doi:10.1109/ICNSURV.2016.7486362.

Boag, C., Neal, A., Loft, S., and Halford, G.S. An Analysis of Relational Complexity in an Air Traffic Control Conflict Detection Task. *Ergonomics*, 49(14):pp. 1508–1526, 2006. doi:10.1080/00140130600779744.

Borghini, G., Aricò, P., DI Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., Golfetti, A., Imbert, J.P., Granger, G., Benhacene, R., Pozzi, S., and Babiloni, F. EEG-Based Cognitive Control Behaviour Assessment: An Ecological study with Professional Air Traffic Controllers. *Scientific Reports*, 7, 2017. doi:10.1038/s41598-017-00633-7.

Borst, C., Bijsterbosch, V.A., van Paassen, M.M., and Mulder, M. Ecological interface design: supporting fault diagnosis of automated advice in a supervisory air traffic control task. *Cognition, Technology & Work*, 19(4):pp. 545–560, 2017. doi:10.1007/s10111-017-0438-y.

Borst, C., Flach, J.M., and Ellerbroek, J. Beyond ecological interface design: Lessons from concerns and misconceptions. *IEEE Transactions on Human-Machine Systems*, 45(2):pp. 164–175, 2015. doi:10.1109/THMS.2014.2364984.

Borst, C., Visser, R.M., van Paassen, M.M., and Mulder, M. Exploring Short-Term Training Effects of Ecological Interfaces: A Case Study in Air Traffic Control. *IEEE Transactions on Human-Machine Systems*, 49(6):pp. 623–632, 2019. doi:10.1109/THMS.2019.2919742.

ten Brink, D.S.A., Klomp, R.E., Borst, C., van Paassen, M.M., and Mulder, M. Flow-based air traffic control: Human-machine interface for steering a path-planning algorithm. In *IEEE International Conference on Systems, Man and Cybernetics*, pp. 3186–3191. 2019. ISBN 9781728145693. doi:10.1109/SMC.2019.8914030.

Calhoun, G. Adaptable (Not Adaptive) Automation: The Forefront of Human–Automation Teaming. *Human Factors*, 2021. doi:10.1177/00187208211037457.

Capiot, K., Korn, B., and Mulder, M. Development of an Algorithm to Determine the Attention Required by Aircraft Under Control Using Fuzzy Logic. In *AIAA Scitech 2019 Forum*. American Institute of Aeronautics and Astronautics, 2019. doi:10.2514/6.2019-1638.

Chen, J.Y.C., Flemisch, F.O., Lyons, J.B., and Neerincx, M.A. Guest editorial: Agent and system transparency. *IEEE Transactions on Human-Machine Systems*, 50(3):pp. 189–193, 2020. doi:10.1109/thms.2020.2988835.

Chiappe, D.L., Vu, K.P.L., and Strybel, T.Z. Situation Awareness in the NextGen Air Traffic Management System. *International Journal of Human-Computer Interaction*, 28(2):pp. 140–151, 2012. doi:10.1080/10447318.2012.634768.

Christoffersen, K. and Woods, D.D. How To Make Automated Systems Team Players. In E. Salas, ed., *Advances in Human Performance and Cognitive Engineering Research*, volume 2, pp. 1–12. Emerald Group Publishing Limited, Leeds, UK, 2002. ISBN 978-0-76230-864-4. doi:10.1016/S1479-3601(02)02003-9.

Cioroaica, E., Buhnova, B., Kuhn, T., and Schneider, D. Building trust in the untrustable. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Society*, ICSE '20, pp. 21–24. ACM, 2020. doi:10.1145/3377815.3381378.

Contissa, G., Sartor, G., Lanzi, P., Marti, P., and Tomasello, P. Liabilities and automation in aviation. In *2nd SESAR Innovation Days*. Braunschweig, Germany, 2012.

Corver, S.C. and Aneziris, O.N. The impact of controller support tools in enroute air traffic control on cognitive error modes: A comparative analysis in two operational environments. *Safety Science*, 71:pp. 2–15, 2015. doi:10.1016/j.ssci.2014.07.018.

Corver, S.C. and Grote, G. Uncertainty Management in Enroute Air Traffic Control: A Field Study Exploring Controller Strategies and Requirements for Automation. *Cognition, Technology & Work*, 18(3):pp. 541–565, 2016. doi:10.1007/s10111-016-0373-3.

Cummings, M.L., Tsonis, C.G., and Cunha, D.C. Complexity Mitigation Through Airspace Structure. In *13th International Symposium on Aviation Psychology*, pp. 159–163. Dayton, OH, USA, 2005.

D'Arcy, J.F. and Della Rocco, P.S. Air Traffic Control Specialist Decision Making and Strategic Planning-A Field Survey. Technical Report DOT/FAA/CT-TN01/05, FAA, 2001. doi:10.1037/e447152006-001.

Dekker, S. On the other side of promise: what should we automate today? In D. Harris, ed., *Human Factors for Civil Flight Deck Design*, chapter 8, pp. 183–198. Routledge, London. UK, 2004. ISBN 9781351929844. doi:10.4324/9781315253039.

Di Flumeri, G., De Crescenzio, F., Berberian, B., Ohneiser, O., Kraemer, J., Aricò, P., Borghini, G., Babiloni, F., Bagassi, S., and Piastra, S. Brain–Computer Interface-Based Adaptive Automation to Prevent Out-Of-The-Loop Phenomenon in Air Traffic Controllers Dealing With Highly Automated Systems. *Frontiers in Human Neuroscience*, 13(296), 2019. doi:10.3389/fnhum.2019.00296.

Dittmann, A., Kallus, K.W., and Van Damme, D. Integrated Task and Job Analysis of Air Traffic Controllers - Phase 3: Baseline Reference of Air Traffic Controller Tasks and Cognitive Processes in the ECAC Area. Technical Report HUM.ET1.ST01.1000-REP-05, Eurocontrol, 2000.

Dow, C. and Histon, J. Enroute ATC Industry Perceptions of Simulation Fidelity. In *18th International Symposium on Aviation Psychology*, pp. 524–529. Dayton, OH, USA, 2015.

Dow, C. and Histon, J.M. An Air Traffic Control Simulation Fidelity Definition and Categorization System. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):pp. 92–96, 2014. doi:10.1177/1541931214581020.

Drew, A. and Makins, N. Initial Study of the Executive Controller and 'Problem-Free' Groups. Technical report, Eurocontrol, 1994.

Duong, V., Gawinowski, G., Nicolaon, J.P., and Smith, D. Sector-Less Air Traffic Management. In *Proceedings of the fourth USA/Europe air traffic management research and development seminar*. Santa Fe, NM, USA, 2001.

Edwards, T.E., Homola, J.R., Mercer, J.S., and Claudatos, L. Multifactor interactions and the air traffic controller: the interaction of situation awareness and workload in association with automation. *Cognition, Technology & Work*, 19(4):pp. 687–698, 2017. doi:10.1007/s10111-017-0445-z.

Edwards, T.E. and Lee, P.U. Towards Designing Graceful Degradation into Trajectory Based Operations: A Human-Machine System Integration Approach. In *17th AIAA Aviation Technology, Integration, and Operations Conference*. American Institute of Aeronautics and Astronautics, Denver, CO, USA, 2017. doi:10.2514/6.2017-4487.

Endsley, M.R. From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, 59(1):pp. 5–27, 2017. doi:10.1177/0018720816681350.

Endsley, M.R. Ironies of artificial intelligence. *Ergonomics*, 66(11):pp. 1656–1668, 2023. doi:10.1080/00140139.2023.2243404.

Endsley, M.R. and Kaber, D.B. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):pp. 462–492, 1999. doi:10.1080/001401399185595.

Endsley, M.R. and Kiris, E.O. The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2):pp. 381–394, 1995. doi:10.1518/001872095779064555.

Enea, G. and Porretta, M. A comparison of 4D-trajectory operations envisioned for NextGen and SESAR, some preliminary findings. In *28th International Congress of the Aeronautical Sciences*. 2012.

Erzberger, H., Lauderdale, T.A., and Chu, Y.C. Automated conflict resolution, arrival management, and weather avoidance for air traffic management. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 226(8):pp. 930–949, 2012. doi:10.1177/0954410011417347.

Etherington, T., Kramer, L., Vie, L.L., Kennedy, K., Bailey, R., and Houston, V. Evaluation of Technology Concepts for Traffic data Management and Relevant audio for Datalink in Commercial Airline Flight Decks. In *2019 Integrated Communications, Navigation and Surveillance Conference*. IEEE, 2019. doi:10.1109/icnsurv.2019.8735135.

Eurocontrol. Eurocontrol Specification for Short Term Conflict Alert. 2007.

Eurocontrol. User Manual for the Base of Aircraft Data (BADA) Revision 3.10. 2012.

Eurocontrol. Eurocontrol Specification for Medium-Term Conflict Detection. 2017.

Eurocontrol. Eurocontrol Seven-Year Forecast 2024-2030 - Autumn 2024 Update. 2024a.

Eurocontrol. LORD Adds Speed and Flexibility to Decisions by Maastricht Controllers. In *Skyway*, volume 79, pp. 54–56. Eurocontrol, 2024b.

Ezzati Amini, R., Katrakazas, C., Riener, A., and Antoniou, C. Interaction of automated driving systems with pedestrians: challenges, current solutions, and recommendations for eHMIs. *Transport Reviews*, 41(6):pp. 788–813, 2021. doi:10.1080/01441647.2021.1914771.

Feigh, K.M. and Pritchett, A.R. Requirements for effective function allocation: A critical review. *Journal of Cognitive Engineering and Decision Making*, 8(1):pp. 23–32, 2014. doi:10.1177/1555343413490945.

Finck, T., Klünker, C.S., and Martins, A. Validation of the Flight Centric ATC Concept using Hungarian Airspace as an Example. In *AIAA AVIATION 2023 Forum*. AIAA, 2023a. doi:10.2514/6.2023-3259.

Finck, T., Klünker, C.S., and Weber, M. Conceptual Analysis of Allocation Strategies for Air Traffic Control Concepts without Conventional Sector Boundaries. In *Integrated Communications, Navigation and Surveillance Conference*. IEEE, 2023b. ISBN 979-8-3503-3362-6. doi:10.1109/ICNS58246.2023.10124289.

Finck, T., Névir, M.C., and Klünker, C.S. Design and Validation of a Flight Centric Workload Model Including ATC Task Change and Considering Influencing Factors. In *33rd Congress of the International Council of the Aeronautical Sciences*. Stockholm (Sweden), 2022.

Finck, T., Temme, A., Tittel, S., and Korn, B. Workload-Based Allocation Strategy for the New Concept of Flight Centric ATC. In *AIAA Aviation Forum and ASCEND*. 2024. doi:10.2514/6.2024-4009.

Fothergill, S. and Neal, A. Conflict-Resolution Heuristics for En Route Air Traffic Management. In *Proc. Human Factors and Ergonomics Society 57th Annual Meeting*, pp. 71–75. San Diego, CA, USA, 2013. ISBN 9780945289432. doi:10.1177/1541931213571018.

Fraga, R.P., Kang, Z., Crutchfield, J.M., and Mandal, S. Visual Search and Conflict Mitigation Strategies Used by Expert en Route Air Traffic Controllers. *Aerospace*, 8(7), 2021. doi:10.3390/aerospace8070170.

Frazzoli, E., Mao, Z.H., Oh, J.H., and Feron, E. Resolution of Conflicts Involving Many Aircraft via Semidefinite Programming. *Journal of Guidance, Control, and Dynamics*, 24(1):pp. 79–86, 2001. doi:10.2514/2.4678.

Gariel, M. and Feron, E. 3D conflict avoidance under uncertainties. In *IEEE/AIAA 28th Digital Avionics Systems Conference*, volume 2, pp. 4.E.3–1–4.E.3–8. IEEE, 2009. doi:10.1109/dasc.2009.5347480.

Gerdes, I., Jameel, M., Hunger, R., Christoffels, L., and Gürlük, H. The Automation Evolves: Concept for a Highly Automated Controller Working Position. In *33rd Congress of the International Council of the Aeronautical Sciences*. Stockholm, Sweden, 2022.

Gerdes, I., Temme, A., and Schultz, M. Dynamic airspace sectorisation for flight-centric operations. *Transportation Research Part C: Emerging Technologies*, 95(January):pp. 460–480, 2018. doi:10.1016/j.trc.2018.07.032.

Gil, P., Vidaller, A., Cano, M., Gómez, D., López, J.A., Ceñal, N., Chaves, J., and Ruiz-Artaza, F. Feasibility and Benefits of Implementing Flight Centric ATC in the Spanish upper airspace. In *13th SESAR Innovation Days*. Seville, Spain, 2023.

Gilbert, G. Historical development of the air traffic control system. *IEEE Transactions on Communications*, 21(5):pp. 364–375, 1973. doi:10.1109/tcom.1973.1091699.

Griffiths, N., Bowden, V., Wee, S., Strickland, L., and Loft, S. Operator selection for human-automation teaming: The role of manual task skill in predicting automation failure intervention. *Applied Ergonomics*, 118:104288, 2024. doi:10.1016/j.apergo.2024.104288.

Gronlund, S.D., Ohrt, D.D., Dougherty, M.R.P., Perry, J.L., and Manning, C.A. Role of memory in air traffic control. *Journal of Experimental Psychology: Applied*, 4(3):pp. 263–280, 1998. doi:10.1037/1076-898X.4.3.263.

Hancock, P.A. Automation: How much is too much? *Ergonomics*, 57(3):pp. 449–454, 2014. doi:10.1080/00140139.2013.816375.

Harrison, J., Izzetoğlu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P.A., Bunce, S.C., and Onaral, B. Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems*, 44(4):pp. 429–440, 2014. doi:10.1109/THMS.2014.2319822.

Haugg, E., Poppe, M., Herr, S., Putz, T., Svoboda, J., and Sosovicka, R. The usability of ADS-C EPP data for air traffic control applications. In *IEEE/AIAA 34th Digital Avionics Systems Conference*. IEEE, 2015. doi:10.1109/dasc.2015.7311364.

Hendrickx, P. Advanced Automation in ATC at Eurocontrol Maastricht UAC. Presentation, 2023. Available at: www.faa.gov/air_traffic/flight_info/aeronav/atiec/media/Presentations/2023/Day_3/Advanced%20Automation.pdf.

Hendrickx, P. and Tisza, A.B. Eurocontrol Maastricht Upper Area Control Centre OPS & Automation Strategy. Presentation, 2019. Available at: skybrary.aero/bookshelf/books/5341.pdf.

Hessels, R.S., Niehorster, D.C., Kemner, C., and Hooge, I.T.C. Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, 49(5):pp. 1802–1823, 2017. doi:10.3758/s13428-016-0822-1.

Hilburn, B. Cognitive Complexity in Air Traffic Control - A Literature Review. Technical report, Eurocontrol, 2004. Available at: eurocontrol.int/sites/default/files/library/007_Cognitive_Complexity_in_ATC.pdf.

Hilburn, B., Westin, C., and Borst, C. Will Controllers Accept a Machine That Thinks like They Think? The Role of Strategic Conformance in Decision Aiding Automation. *Air Traffic Control Quarterly*, 22(2):pp. 115–136, 2014. doi:10.2514/atcq.22.2.115.

Hoc, J.M. and Carlier, X. Role of a Common Frame of Reference in Cognitive Cooperation: Sharing Tasks between Agents in Air Traffic Control. *Cognition, Technology & Work*, 4(1):pp. 37–47, 2002. doi:10.1007/s101110200003.

Hoekstra, J.M., Maas, J.B., Tra, M.A.P., and Sunil, E. How Do Layered Airspace Design Parameters Affect Airspace Capacity and Safety? In *7th International Conference on Research in Air Transportation*. Philadelphia, PA, USA, 2016.

Holford, W.D. An Ethical Inquiry of the Effect of Cockpit Automation on the Responsibilities of Airline Pilots: Dissonance or Meaningful Control? *Journal of Business Ethics*, 176(1):pp. 141–157, 2020. doi:10.1007/s10551-020-04640-z.

Hopkin, V.D. The Impact of Automation on Air Traffic Control Specialists. In M.W. Smolensky and E.S. Stein, eds., *Human Factors in Air Traffic Control*, chapter 13, pp. 391–419. Academic Press, San Diego, 1998. ISBN 0126530106.

Hunger, R., Christoffels, L., Friedrich, M., Jameel, M., Pick, A., Gerdes, I., von der Nahmer, P.M., and Sobotzki, F. Lesson Learned: Design and Perception of Single Controller Operations Support Tools. In *International Conference on Human-Computer Interaction*, pp. 15–33. Springer Nature Switzerland, 2024. ISBN 9783031607318. doi:10.1007/978-3-031-60731-8_2.

Hunt, V.R. and Zellweger, A. The FAA's Advanced Automation System: Strategies for Future Air Traffic Control Systems. *Computer*, 20(2):pp. 19–32, 1987. doi:10.1109/MC.1987.1663475.

IJtsma, M., Borst, C., van Paassen, M.M., and Mulder, M. Evaluation of a Decision-Based Invocation Strategy for Adaptive Support for Air Traffic Control. *IEEE Transactions on Human-Machine Systems*, 52(6):pp. 1135–1146, 2022. doi:10.1109/THMS.2022.3208817.

International Civil Aviation Organization. *Global Operational Data Link (GOLD) Manual (Doc 10037)*. ICAO, Montreal, QC, Canada, first edition, 2017.

Isufaj, R., Omeri, M., Piera, M.A., Saez Valls, J., and Verdonk Gallego, C.E. From Single Aircraft to Communities: A Neutral Interpretation of Air Traffic Complexity Dynamics. *Aerospace*, 9(10), 2022. doi:10.3390/aerospace9100613.

Jagasits, V. MUAC Pioneers Digital Connectivity Between Ground and Air with ADS-C. In *Skyway*, volume 80, pp. 56–58. EUROCONTROL, 2024.

Jamieson, G.A., Skraaning, G., and Joe, J. The B737 MAX 8 Accidents as Operational Experiences With Automation Transparency. *IEEE Transactions on Human-Machine Systems*, 52(4):pp. 794–797, 2022. doi:10.1109/thms.2022.3164774.

Jans, M., Borst, C., van Paassen, M.M., and Mulder, M. Effect of ATC Automation Transparency on Acceptance of Resolution Advisories. In *14th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems*, pp. 353–358. Tallinn, Estonia, 2019. doi:10.1016/j.ifacol.2019.12.087.

Janssen, C.P., Donker, S.F., Brumby, D.P., and Kun, A.L. History and future of human-automation interaction. *International Journal of Human Computer Studies*, 131(May):pp. 99–107, 2019. doi:10.1016/j.ijhcs.2019.05.006.

Jasek, M., Pioch, N., and Zeltzer, D. Enhanced Visual Displays for Air Traffic Control Collision Prediction. In *6th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, pp. 553–558. Cambridge, MA, USA, 1995. doi:10.1016/s1474-6670(17)45290-6.

Jha, P.D., Bisantz, A.M., Parasuraman, R., and Drury, C.G. Air traffic controllers' performance in advance air traffic management system: Part I-performance results. *International Journal of Aviation Psychology*, 21(3):pp. 283–305, 2011. doi:10.1080/10508414.2011.582456.

Joe, J.C., O'Hara, J., Medema, H.D., and Oxstrand, J.H. Identifying requirements for effective human-automation teamwork. *PSAM 2014 - Probabilistic Safety Assessment and Management*, 2014.

Joint Planning and Development Office. Concept of Operations for the Next Generation Air Transportation System. Technical report, FAA, 2011.

Kaber, D.B. Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation. *Journal of Cognitive Engineering and Decision Making*, 12(1):pp. 7–24, 2018. doi:10.1177/1555343417737203.

Kaber, D.B. and Endsley, M.R. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2):pp. 113–153, 2004. doi:10.1080/1463922021000054335.

Kaber, D.B., Wright, M.C., Prinzel III, L.J., and Clamann, M.P. Adaptive Automation of Human-Machine System Information-Processing Functions. *Human Factors*, 47(4):pp. 730–741, 2005. doi:10.1518/001872005775570989.

Kallus, K.W., Van Damme, D., and Dittmann, A. Integrated Task and Job Analysis of Air Traffic Controllers - Phase 2: Task Analysis of En-route Controllers. Technical Report HUM.ET1.ST01.1000-REP-04, Eurocontrol, 1999.

Kang, Z. and Landry, S.J. An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering. *IEEE Transactions on Human-Machine Systems*, 45(1):pp. 13–24, 2015. doi:10.1109/THMS.2014.2363121.

Kannan, R. and Lasky, R.C. Autonomous Vehicles Still Decades Away: 2019. In *Pan Pacific Microelectronics Symposium*. IEEE, 2020. doi:10.23919/panpacific48324.2020.9059394.

Kassner, M., Patera, W., and Bulling, A. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 1151–1160. ACM, Seattle, Washington, 2014. ISBN 9781450330473. doi:10.1145/2638728.2641695.

Kenny, S. and Li, W.C. The Effect of the COVID-19 Pandemic on Air Traffic Controllers' Performance: Addressing Skill Fade Concerns in Preparation for the Return to Normal Operations. In *International Conference on Human-Computer Interaction*, pp. 300–313. 2022. ISBN 9783031060861. doi:10.1007/978-3-031-06086-1_23.

Khastgir, S., Birrell, S., Dhadyalla, G., and Jennings, P. Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles. *Transportation Research Part C: Emerging Technologies*, 96:pp. 290–303, 2018. doi:10.1016/j.trc.2018.07.001.

Kim, M., Borst, C., and Mulder, M. Situation Awareness Prompts: Bridging the Gap between Supervisory and Manual Air Traffic Control. In *15th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems*, pp. 13–18. San José, CA, USA, 2022. doi:10.1016/j.ifacol.2022.10.224.

Kirlik, A. Modeling strategic behavior in human-automation interaction: Why an 'aid' can (and should) go unused. *Human Factors*, 35(2):pp. 221–242, 1993. doi:10.1177/001872089303500203.

Kirwan, B. and Flynn, M. Investigating Air Traffic Controller Conflict Resolution Strategies. Technical Report ASA.01.CORA.2.DEL04-B.RS, Eurocontrol, Brussels, Belgium, 2002.

Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., and Feltovich, P.J. Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, 19(6):pp. 91–95, 2004. doi:10.1109/MIS.2004.74.

Klomp, R.E., Borst, C., van Paassen, M.M., and Mulder, M. Expertise Level, Control Strategies, and Robustness in Future Air Traffic Control Decision Aiding. *IEEE Transactions on Human-Machine Systems*, 46(2):pp. 255–266, 2016. doi:10.1109/THMS.2015.2417535.

Klomp, R.E., Riegman, R., Borst, C., Mulder, M., and van Paassen, M.M. Solution space concept: Human-machine interface for 4D trajectory management. In *13th USA/Europe Air Traffic Management Research and Development Seminar (ATM2019)*. 2019.

Klünker, C.S., Weber, M., Finck, T., and Taraldsson, Å. Effects on ATCO Training when Working in a Flight Centric ATC Environment. In *IEEE/AIAA 42nd Digital Avionics Systems Conference*. IEEE, Barcelona, Spain, 2023. doi:10.1109/DASC58513.2023.10311161.

Korn, B., Mollwitz, V., Finck, T., Edinger, C., and Tittel, S. Validating Sectorless ATM in the Hungarian Airspace: Results of Human in the Loop Simulations. In *Integrated Communications, Navigation and Surveillance Conference*. IEEE, Herndon, VA, USA, 2020. ISBN 9781728172705. doi:10.1109/icns50378.2020.9223018.

Kumbhar, A.K.V., Lyu, W., and Borst, C. Determining Flight Complexity and Relevance: Flight-Centric Filtering for Air Traffic Control. In *14th SESAR Innovation Days*. Rome, Italy, 2024.

Lagu, A.V. and Landry, S.J. Roadmap for the next generation of dynamic function allocation theories and strategies. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 21(1):pp. 14–28, 2010. doi:10.1002/hfm.20209.

Landry, S.J. Flight Deck Automation. In S.Y. Nof, ed., *Springer Handbook of Automation*, chapter 68, pp. 1215–1239. Springer Berlin Heidelberg, 2009. ISBN 9783540788317. doi:10.1007/978-3-540-78831-7_68.

Landry, S.J. The next generation air transportation system: An approach to function allocation. *Human Factors and Ergonomics In Manufacturing*, 22(4):pp. 351–361, 2012. doi:10.1002/hfm.20265.

Langford, K., Kille, T., Lee, S.Y., Zhang, Y., and Bates, P.R. “In Automation We Trust” - Australian Air Traffic Controller Perspectives of Increasing Automation in Air Traffic Management. *Transport Policy*, 125:pp. 352–362, 2022. doi:10.1016/j.tranpol.2022.07.001.

Lanzi, P., Taurino, D., Contissa, G., and Sartor, G. ARGOS Legal Case: Report of Results. Technical report, Deep Blue, 2021.

Lee, J.D. and See, K.A. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1):pp. 50–80, 2004. doi:10.1518/hfes.46.1.50_30392.

Lee, K., Feron, E., and Pritchett, A.R. Describing Airspace Complexity: Airspace Response to Disturbances. *Journal of Guidance, Control, and Dynamics*, 32(1):pp. 210–222, 2009. doi:10.2514/1.36308.

Loft, S., Sanderson, P.M., Neal, A., and Mooij, M. Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications. *Human Factors*, 49(3):pp. 376–399, 2007. doi:10.1518/001872007X197017.

Lyons, J.B., Sycara, K., Lewis, M., and Capiola, A. Human–Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in Psychology*, 12:589585, 2021. doi:10.3389/fpsyg.2021.589585.

Major, L.M. and Hansman, R.J. Human-Centered Systems Analysis of Mixed Equipage in Oceanic Air Traffic Control. Technical Report ICAT-2004-6, MIT International Center for Air Transportation, Cambridge, MA, USA, 2004.

Martin, L., Bienert, N., Claudatos, L., Gujral, V., Kraut, J., and Mercer, J. Effects of task allocation on air traffic management human-automation system performance. In *IEEE/AIAA 35th Digital Avionics Systems Conference*. IEEE, 2016. doi:10.1109/dasc.2016.7777996.

Martins, A.P.G., Finck, T., Mollwitz, V., Kling, F., and Rohács, D. Feasibility study of Flight Centric Mode of Operations. In *9th SESAR Innovation Days*. Athens, Greece, 2019.

Martins, A.P.G., Kluenker, C.S., Edinger, C., Tittel, S., and Weber, M. Evaluation of the Impact on Air Traffic Controllers of Air-Ground Datalink Technologies (LDACS and Current CPDLC) in Flight Centric ATC. In *AIAA DATC/IEEE 43rd Digital Avionics Systems Conference*, pp. 1–9. IEEE, 2024. doi:10.1109/dasc62030.2024.10749213.

Maynard, P.W. and Rantanen, E.M. Teams, Teamwork, and Automation in Air Traffic Control. In *13th International Symposium on Aviation Psychology*, pp. 484–489. Dayton, OH, USA, 2005.

McDonnell, A.S., Crabtree, K.W., Cooper, J.M., and Strayer, D.L. This is your brain on autopilot 2.0: The influence of practice on driver workload and engagement during on-road, partially automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 66(8):pp. 2025–2040, 2023. doi:10.1177/00187208231201054.

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J.D., Johansson, E., Green, C.A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., and Keinath, A. The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*, 21(1):pp. 87–98, 2019. doi:10.1007/s10111-018-0525-8.

Mercer, J.S., Gabets, C., Gomez, A.N., Edwards, T.E., Bienert, N., Claudatos, L., and Homola, J.R. How Important is Conflict Detection to the Conflict Resolution Task? In N.A. Stanton, S.J. Landry, G. Di Bucchianico, and A. Vallicelli, eds., *Advances in Human Aspects of Transportation*, pp. 103–115. Springer, 2017. ISBN 9783319416816. doi:10.1007/978-3-319-41682-3_9.

van de Merwe, K., Mallam, S., and Nazir, S. Agent transparency, situation awareness, mental workload, and operator performance: A systematic literature review. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 66(1):pp. 180–208, 2024. doi:10.1177/00187208221077804.

Metzger, U. and Parasuraman, R. The Role of the Air Traffic Controller in Future Air Traffic Management: An Empirical Study of Active Control Versus Passive Monitoring. *Human Factors*, 43(4):pp. 519–528, 2001. doi:10.1518/001872001775870421.

Metzger, U. and Parasuraman, R. Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload. *Human Factors*, 47(1):pp. 35–49, 2005. doi:10.1518/0018720053653802.

Metzger, U. and Parasuraman, R. Effects of automated conflict cuing and traffic density on air traffic controller performance and visual attention in a datalink environment. *International Journal of Aviation Psychology*, 16(4):pp. 343–362, 2006. doi:10.1207/s15327108ijap1604_1.

Millot, P. and Mandiau, R. Man-Machine Cooperative Organizations: Formal and Pragmatic Implementation Methods. In J.M. Hoc, P.C. Cacciabue, and E. Hollnagel, eds., *Expertise and Technology: Cognition & Human-Computer Cooperation*, chapter 13, pp. 213–228. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, USA, 1995.

Mogford, R.H., Guttman, J.A., Morrow, S.L., and Kopardekar, P. The Complexity Construct in Air Traffic Control: A Review and Synthesis of the Literature. Technical Report DOT/FAA/CT-TN95/22, FAA, 1995.

Moriarty, D. *Automation management*, chapter 9, pp. 243–268. Academic Press, Cambrigde, MA, USA, 2015. ISBN 9780124202443. doi:10.1016/B978-0-12-420244-3.00009-1.

Niessen, C. and Eyferth, K. A Model of the Air Traffic Controller's Picture. *Safety Science*, 37:pp. 187–202, 2001. doi:10.1016/S0925-7535(00)00048-5.

Nolan, M.S. *Fundamentals of Air Traffic Control*. Delmar Cengage Learning, fifth edition, 2011. ISBN 9781435482722.

Nolan, M.S. The Design of the Air Traffic Control System. In S.J. Landry, ed., *Handbook of Human Factors in Air Transportation Systems*, chapter 3, pp. 31–50. CRC Press, Boca Raton, FL, USA, 2017. ISBN 9781315116549. doi:10.1201/9781315116549-3.

Nordhoff, S., Lee, J.D., Calvert, S.C., Berge, S., Hagenzieker, M., and Happee, R. (Mis-)use of standard Autopilot and Full Self-Driving (FSD) Beta: Results from interviews with users of Tesla's FSD Beta. *Frontiers in Psychology*, 14, 2023. doi:10.3389/fpsyg.2023.1101520.

Nordman, A., Meyer, L., Klang, K.J., Lundberg, J., and Vrotsou, K. Extraction of CD&R Work Phases from Eye-Tracking and Simulator Logs: A Topic Modelling Approach. *Aerospace*, 10(7):p. 595, 2023. doi:10.3390/aerospace10070595.

Norman, D.A. The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 327(1241):pp. 585–593, 1990. doi:10.1098/rstb.1990.0101.

Norman, D.A. The Human Side of Automation. In G. Meyer and S. Beiker, eds., *Road Vehicle Automation 2*, Lecture Notes in Mobility, chapter 7, pp. 73–79. Springer International Publishing, 2015. ISBN 9783319190785. doi:10.1007/978-3-319-19078-5_7.

Nyholm, S. and Smids, J. Automated cars meet human drivers: responsible human-robot coordination and the ethics of mixed traffic. *Ethics and Information Technology*, 22:pp. 335–344, 2020. doi:10.1007/s10676-018-9445-9.

O'Neill, T.A., Flathmann, C., McNeese, N.J., and Salas, E. Human-autonomy teaming: Need for a guiding team-based framework? *Computers in Human Behavior*, 146:p. 107762, 2023. doi:10.1016/j.chb.2023.107762.

O'Neill, T.A., McNeese, N.J., Barron, A., and Schelble, B. Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors*, 64(5):pp. 904–938, 2022. doi:10.1177/0018720820960865.

Onnasch, L., Wickens, C.D., Li, H., and Manzey, D.H. Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56(3):pp. 476–488, 2014. doi:10.1177/0018720813501549.

Ortner, P., Steinhöfler, R., Leitgeb, E., and Flühr, H. Augmented Air Traffic Control System — Artificial Intelligence as Digital Assistance System to Predict Air Traffic Conflicts. *AI*, 3(3):pp. 623–644, 2022. doi:10.3390/ai3030036.

van Paassen, M.M., Borst, C., Mulder, M., de Rooij, G., Dijkstra, F., and Tisza, A.B. On the Appropriate Participant Expertise for Display Evaluation Studies. In *22nd International Symposium on Aviation Psychology*. Rochester, NY, USA, 2023.

Pacaux-Lemoine, M.P. and Flemisch, F.O. Layers of shared and cooperative control, assistance, and automation. *Cognition, Technology & Work*, 21(4):pp. 579–591, 2019. doi:10.1007/s10111-018-0537-4.

Paglione, M., Young, C.M., Torres, S., Hochwarth, J.K., McDonald, G., Bronsvoort, J., and Boucquey, J. Operational Impact of Trajectory Prediction Accuracy on Air Traffic Automation Tools. In *IEEE/AIAA 36th Digital Avionics Systems Conference*. St. Petersburg, FL, USA, 2017. doi:10.1109/DASC.2017.8102050.

Parasuraman, R. and Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2):pp. 230–253, 1997. doi:10.1518/001872097778543886.

Parasuraman, R., Sheridan, T.B., and Wickens, C.D. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, 30(3):pp. 286–297, 2000. doi:10.1109/3468.844354.

Parasuraman, R. and Wickens, C.D. Humans: Still vital after all these years of automation. *Human Factors*, 50(3):pp. 511–520, 2008. doi:10.1518/001872008X312198.

Pavlović, G., Jovanović, R., and Stanojević, M. Rostering of air traffic controllers – current practices in Europe, key performance trade-offs and future opportunities. *Journal of Air Transport Management*, 113:102472, 2023. doi:10.1016/j.jairtraman.2023.102472.

Pepper, N., Thomas, M., De Ath, G., Olivier, E., Cannon, R., Everson, R., and Dodwell, T. A probabilistic model for aircraft in climb using monotonic functional Gaussian process emulators. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 479(2271), 2023. doi:10.1098/rspa.2022.0607.

Pérez-Castán, J.A., Pérez-Sanz, L., Bowen-Varela, J., Serrano-Mira, L., Radišić, T., and Feuerle, T. Machine Learning Classification Techniques Applied to Static Air Traffic Conflict Detection. In *IOP Conference Series: Materials Science and Engineering*, volume 1226. 2022. doi:10.1088/1757-899x/1226/1/012019.

Pérez Moreno, F., Gómez Comendador, V.F., Delgado-Aguilera Jurado, R., Zamarreño Suárez, M., Janisch, D., and Arnaldo Valdés, R.M. Determination of Air Traffic Complexity Most Influential Parameters Based on Machine Learning Models. *Symmetry*, 14(12):p. 2629, 2022. doi:10.3390/sym14122629.

Pfeiffer, L., Müller, N.H., and Rosenthal, P. A survey of visual and interactive methods for air traffic control data. In *19th International Conference on Information Visualisation*, pp. 574–577. IEEE, 2015. ISBN 9781467375689. doi:10.1109/iV.2015.102.

Pham, D.T., Tran, P.N., Alam, S., Duong, V., and Delahaye, D. Deep reinforcement learning based path stretch vector resolution in dense traffic with uncertainties. *Transportation Research Part C: Emerging Technologies*, 135:p. 103463, 2022. doi:10.1016/j.trc.2021.103463.

Pop, V.L., Stearman, E.J., Kazi, S., and Durso, F.T. Using Engagement to Negate Vigilance Decrements in the NextGen Environment. *International Journal of Human-Computer Interaction*, 28(2):pp. 99–106, 2012. doi:10.1080/10447318.2012.634759.

Prandini, M., Piroddi, L., Puechmorel, S., and Brázdilová, S.L. Toward Air Traffic Complexity Assessment in New Generation Air Traffic Management Systems.

IEEE Transactions on Intelligent Transportation Systems, 12(3):pp. 809–818, 2011. doi:10.1109/TITS.2011.2113175.

Prevot, T., Homola, J.R., Martin, L.H., Mercer, J.S., and Cabrall, C.D.D. Toward Automated Air Traffic Control - Investigating a Fundamental Paradigm Shift in Human/Systems Interaction. *International Journal of Human-Computer Interaction*, 28(2):pp. 77–98, 2012. doi:10.1080/10447318.2012.634756.

Prevot, T. and Lee, P.U. Trajectory-Based Complexity (TBX): A modified aircraft count to predict sector complexity during trajectory-based operations. In *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*, pp. 3A3–1–3A3–15. IEEE, 2011. doi:10.1109/dasc.2011.6096045.

Prevot, T., Lee, P.U., Smith, N., and Palmer, E. ATC Technologies for Controller-Managed and Autonomous Flight Operations. In *AIAA Guidance, Navigation, and Control Conference*. San Francisco, CA, USA, 2005. ISBN 1563477378. doi:10.2514/6.2005-6043.

Pritchett, A.R., Kim, S.Y., and Feigh, K.M. Measuring human-automation function allocation. *Journal of Cognitive Engineering and Decision Making*, 8(1):pp. 52–77, 2014. doi:10.1177/1555343413490166.

Pérez-Castán, J.A., Pérez-Sanz, L., Serrano-Mira, L., Saéz-Hernando, F.J., Rodríguez Gauxachs, I., and Gómez-Comendador, V.F. Design of an ATC Tool for Conflict Detection Based on Machine Learning Techniques. *Aerospace*, 9(2):p. 67, 2022. doi:10.3390/aerospace9020067.

Pérez-Castán, J.A., Rodríguez-Sanz, ☒., Pérez Sanz, L., Arnaldo Valdés, R.M., Gómez Comendador, V.F., Greatti, C., and Serrano-Mira, L. Probabilistic Strategic Conflict-Management for 4D Trajectories in Free-Route Airspace. *Entropy*, 22(2):p. 159, 2020. doi:10.3390/e22020159.

Rantanen, E.M. and Nunes, A. Hierarchical Conflict Detection in Air Traffic Control. *The International Journal of Aviation Psychology*, 15(4):pp. 339–362, 2005. doi:10.1207/s15327108ijap1504_3.

Rasmussen, J. Skills, Rules, and Knowledge; Signals, Signs,and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3):pp. 257–266, 1983. doi:10.1109/TSMC.1983.6313160.

Rasmussen, J. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. Elsevier Science Inc., New York, NY, USA, 1986. doi:10.1016/0306-4573(88)90080-5.

Reek, S. *Real-Time Label Overlap Avoidance for Air Traffic Controllers using Probabilistic Roadmaps*. Master's thesis, Utrecht University, 2010.

Renner, P., Rohács, D., Papp, G., and Kling, F. The Effects of the Introduction of Free Route (HUFRA, Hungarian Free Route Airspace) in the Hungarian Airspace. In *8th SESAR Innovation Days*. Salzburg, Austria, 2018.

Rieth, M. and Hagemann, V. Automation as an equal team player for humans? – A view into the field and implications for research and practice. *Applied Ergonomics*, 98:103552, 2022. doi:10.1016/j.apergo.2021.103552.

Rieth, M., Onnasch, L., and Hagemann, V. Adaptable automation for a more human-centered work design? effects on human perception and behavior. *International Journal of Human-Computer Studies*, 186:103246, 2024. doi:10.1016/j.ijhcs.2024.103246.

de Rooij, G., Borst, C., van Paassen, M.M., and Mulder, M. Flight Allocation in Shared Human-Automation En-Route Air Traffic Control. In *21st International Symposium on Aviation Psychology*, pp. 172–177. Online, 2021.

de Rooij, G., Borst, C., van Paassen, M.M., and Mulder, M. Preventing Scenario Recognition in Human-in-the-loop Air Traffic Control Research. In *22nd International Symposium on Aviation Psychology*. Rochester, NY, USA, 2023a.

de Rooij, G., Stienstra, A., Borst, C., Tisza, A.B., van Paassen, M.M., and Mulder, M. Contributing Factors to Flight-Centric Complexity in En-Route Air Traffic Control. In *Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)*. 2023b.

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

de Rooij, G., Tisza, A.B., Borst, C., van Paassen, M.M., and Mulder, M. Towards Human-Automation Teamwork in Shared En-Route Air Traffic Control: Task Analysis. In *IEEE International Conference on Human-Machine Systems*. 2022. ISBN 9781665452380. doi:10.1109/ICHMS56717.2022.9980715.

Rovira, E. and Parasuraman, R. Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human Factors*, 52(3):pp. 411–425, 2010. doi:10.1177/0018720810375692.

Rødseth, Ø.J. Constrained Autonomy for a Better Human–Automation Interface. In S.O. Johnsen and T. Porathe, eds., *Sensemaking in Safety Critical and Complex Situations*, chapter 14, pp. 235–247. CRC Press, Boca Raton, FL, USA, 2021. ISBN 9781003003816. doi:10.1201/9781003003816-14.

SAE International. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. 2021. doi:10.4271/J3016_202104. Ground Vehicle Standard J3016_202104.

Salas, E., Sims, D.E., and Shawn Burke, C. Is there A ”big five” in teamwork? *Small Group Research*, 36(5):pp. 555–599, 2005. doi:10.1177/1046496405277134.

Sarter, N.B. and Woods, D.D. How in the World Did We Ever Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):pp. 5–19, 1995. doi:10.1518/001872095779049516.

Save, L. and Feuerberg, B. Designing Human-Automation Interaction: a new level of Automation Taxonomy. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier, eds., *Human Factors: a view from an integrative perspective*. 2012.

Schmitt, A.R., Edinger, C., and Korn, B. Balancing Controller Workload Within a Sectorless ATM Concept. *CEAS Aeronautical Journal*, 2:pp. 35–41, 2011. doi:10.1007/s13272-011-0031-7.

Schricker, B.C., Franceschini, R.W., and Johnson, T.C. Fidelity evaluation framework. In *Proceedings. 34th Annual Simulation Symposium*, SIMSYM-01, pp. 109–116. IEEE Comput. Soc, 2001. doi:10.1109/simsym.2001.922122.

Schultz, M., Tominaga, K., Itoh, E., and Duong, V.N. Introduction of moving sectors for flow-centric airspace management. In *13th SESAR Innovation Days*. 2023. doi:10.61009/SID.2023.1.35.

Schutte, P.C. How to make the most of your human: design considerations for human–machine interactions. *Cognition, Technology & Work*, 19(2-3):pp. 233–249, 2017. doi:10.1007/s10111-017-0418-2.

See, J.E. Human Readiness Levels Explained. *Ergonomics in Design*, 29(4):pp. 5–10, 2021. doi:10.1177/10648046211017410.

SESAR Joint Undertaking. A proposal for the future architecture of the European airspace. 2019. doi:10.2829/309090.

SESAR Joint Undertaking. European ATM Master Plan 2020. 2020. doi:10.2829/695700.

SESAR Joint Undertaking. European ATM Master Plan 2025. 2024. doi:10.2829/9855255.

Sgorcea, R., Symionow, W., Balakrishna, M., and Becher, T. Integrating avionics standards with ground air traffic management systems to meet 2025 operational needs. In *IEEE/AIAA 35th Digital Avionics Systems Conference*. IEEE, 2016. doi:10.1109/dasc.2016.7778060.

Sheridan, T.B. and Parasuraman, R. Human-Automation Interaction. *Reviews of Human Factors and Ergonomics*, 1(1):pp. 89–129, 2005. doi:10.1518/155723405783703082.

Sheridan, T.B. and Verplank, W.L. Human and Computer Control of Undersea Teleoperators. Technical report, Massachussetts Institute of Technology, Cambridge, MA, USA, 1978.

Sollenberger, R.L. and Hale, M. Human-in-the-Loop Investigation of Variable Separation Standards in the En Route Air Traffic Control Environment. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*, pp. 66–70. Las Vegas, NV, USA, 2011. ISBN 9780945289395. doi:10.1177/1071181311551014.

Soraji, Y., Furuta, K., Kanno, T., Aoyama, H., Inoue, S., Karikawa, D., and Takahashi, M. Cognitive model of team cooperation in en-route air traffic control. *Cognition, Technology & Work*, 14(2):pp. 93–105, 2012. doi:10.1007/s10111-010-0168-x.

Springer, A. and Whittaker, S. Progressive Disclosure: When, Why, and How Do Users Want Algorithmic Transparency Information? *ACM Transactions on Interactive Intelligent Systems*, 10(4):pp. 1–32, 2020. doi:10.1145/3374218.

Stayton, E. and Stilgoe, J. It's Time to Rethink Levels of Automation for Self-Driving Vehicles [Opinion]. *IEEE Technology and Society Magazine*, 39(3):pp. 13–19, 2020. doi:10.1109/MTS.2020.3012315.

Strauch, B. Ironies of Automation: Still Unresolved After All These Years. *IEEE Transactions on Human-Machine Systems*, 48(5):pp. 419–433, 2018. doi:10.1109/THMS.2017.2732506.

Strybel, T.Z., Vu, K.P.L., Chiappe, D.L., Morgan, C.A., Morales, G., and Battiste, V. Effects of NextGen Concepts of Operation for Separation Assurance and Interval Management on Air Traffic Controller Situation Awareness, Workload, and Performance. *International Journal of Aviation Psychology*, 26(1-2):pp. 1–14, 2016. doi:10.1080/10508414.2016.1235363.

Svensson, Å., Ohlander, U., and Lundberg, J. Design implications for teamwork in ATC. *Cognition, Technology & Work*, 22:pp. 409–426, 2019. doi:10.1007/s10111-019-00579-y.

Tattersall, A.J. and Foord, P.S. An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5):pp. 740–748, 1996. doi:10.1080/00140139608964495.

Trapsilawati, F., Chen, C.H., Wickens, C.D., and Qu, X. Integration of conflict resolution automation and vertical situation display for on-ground air traffic control operations. *Journal of Navigation*, 74(3):pp. 619–632, 2021. doi:10.1017/S0373463320000703.

Vagia, M., Transeth, A.A., and Fjerdingen, S.A. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, 53:pp. 190–202, 2016. doi:10.1016/j.apergo.2015.09.013.

Vanderhaegen, F., Crevits, I., Debernard, S., and Mi, P. Human-Machine Cooperation: Toward an Activity Regulation Assistance for Different Air Traffic Control Levels. *International Journal of Human-Computer Interaction*, 6(1):pp. 65–104, 1994. doi:10.1080/10447319409526084.

de Visser, E.J., Pak, R., and Shaw, T.H. From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction. *Ergonomics*, 61(10):pp. 1409–1427, 2018. doi:10.1080/00140139.2018.1457725.

Volf, P. Comparison of the Flight Centric and Conventional Air Traffic Control. In *Integrated Communications, Navigation and Surveillance Conference*. IEEE, 2019. ISBN 9781728118932. doi:10.1109/ICNSURV.2019.8735316.

Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908(134):pp. 198–287, 1908. doi:10.1515/crll.1908.134.198.

Wasserstein, R.L., Schirm, A.L., and Lazar, N.A. Moving to a World Beyond "p < 0.05". *The American Statistician*, 73(S1):pp. 1–19, 2019. doi:10.1080/00031305.2019.1583913. Editorial.

Westin, C. *Strategic Conformance: Exploring Acceptance of Individual-Sensitive Automation for Air Traffic Control*. PhD thesis, Delft University of Technology, 2017. doi:10.4233/uuid:49c6fe9d-2d29-420a-91a2-a97e2049e15e.

Westin, C., Borst, C., and Hilburn, B. Automation Transparency and Personalized Decision Support: Air Traffic Controller Interaction with a Resolution Advisory System. In *13th IFAC Symposium on Analysis, Design and Evaluation of Human-Machine Systems*, pp. 201–206. Kyoto, Japan, 2016a. doi:10.1016/j.ifacol.2016.10.520.

Westin, C., Borst, C., and Hilburn, B. Strategic Conformance: Overcoming Acceptance Issues of Decision Aiding Automation? *IEEE Transactions on Human-Machine Systems*, 46(1):pp. 41–52, 2016b. doi:10.1109/THMS.2015.2482480.

Westin, C., Borst, C., van Kampen, E.J., Monteiro, T.N., Boonsong, S., Hilburn, B., Cocchioni, M., and Bonelli, S. Personalized and Transparent AI Support for ATC Conflict Detection and Resolution: an Empirical Study. In *Proceedings of the 12th SESAR Innovation Days*. Budapest, Hungary, 2022.

Wickens, C.D. Designing for Situation Awareness and Trust in Automation. In *IFAC Conference on Integrated Systems Engineering*, pp. 365–370. Baden-Baden, Germany, 1995. doi:10.1016/s1474-6670(17)46646-8.

Wickens, C.D. The Psychology of Aviation Surprise: An 8 Year Update Regarding the Noticing of Black Swans. In *15th International Symposium on Aviation Psychology*. Dayton, OH, USA, 2009.

Wickens, C.D. Automation Stages & Levels, 20 Years After. *Journal of Cognitive Engineering and Decision Making*, 12(1):pp. 35–41, 2018. doi:10.1177/1555343417727438.

Wickens, C.D., Clegg, B.A., Vieane, A.Z., and Sebok, A.L. Complacency and Automation Bias in the Use of Imperfect Automation. *Human Factors*, 57(5):pp. 728–739, 2015. doi:10.1177/0018720815581940.

Wilson, I.A.B. and Fleming, K. Controller reactions to free flight in a complex transition sector re-visited using ADS-B+ [ATC]. In *21st Digital Avionics Systems Conference*, pp. 2B5–1 – 2B5–10. IEEE, Irvine, CA, USA, 2002. ISBN 0-7803-7367-7. doi:10.1109/DASC.2002.1067917.

Wright, J.L., Chen, J.Y.C., Barnes, M.J., and Hancock, P.A. Agent Reasoning Transparency's Effect on Operator Workload. In *Proc. Human Factors and Ergonomics Society 60th Annual Meeting*, pp. 249–253. Washington, DC, USA, 2016. doi:10.1177/1541931213601057.

Young, M.S. and Stanton, N.A. To automate or not to automate: advocating the 'cliff-edge' principle. *Ergonomics*, 66(11):pp. 1695–1701, 2023. doi:10.1080/00140139.2023.2270786.

Zon, G.D.R., Roessingh, J.J.M., and Mellett, U. Team Co-ordination Study. Technical report, Eurocontrol, Brussels, Belgium, 2009.

Acronyms

ACC	Area control center
ADS-C	Automatic dependent surveillance - contract
AFL	Actual flight level
AI	Artificial intelligence
ANSP	Air navigation service provider
ARGOS	ATC Real Groundbreaking Operational System
ATC	Air traffic control
ATCO	Air traffic control officer
ATM	Air traffic management
BADA	Base of aircraft data
CC	Coordinating controller
CD&R	Conflict detection and resolution
CFL	Cleared flight level
CONOPS	Concept of operations
COP	Coordination point
CPA	Closest point of approach
CPDLC	Controller-pilot data link communications
CWP	Controller working position
DCT	Direct routing
DECO	Delta and Coastal
EC	Executive controller
EPP	Extended projected profiles
FCA	Flight-centric ATC
FIM	Flight information management
FMS	Flight management system
FOI	Flight of interest
FRA	Free route airspace
HAT	Human-automation team
HMI	Human-machine interface
LOA	Level of automation
LORD	Lateral obstacle and resolution display
LOS	Loss of separation
LVNL	Luchtverkeersleiding Nederland (Air Traffic Control the Netherlands)
MTCD	Medium-term conflict detection
MUAC	Maastricht Upper Area Control Centre
NCOP	Entry coordination point

NFL	Entry flight level
NTCA	Near-term conflict alert
PVD	Plan view display
R/T	Radiotelephony
RPS	Radar position symbol
SA	Situation awareness
SESAR	Single European Sky ATM Research
SRK	Skill, rule and knowledge
STCA	Short-term conflict alert
TBO	Trajectory-based operations
TFL	Transfer flight level
VERA	Verification and advice tool
XCOP	Exit coordination point

Acknowledgments

Digital version intentionally left blank.

Curriculum vitae

Gijs de Rooij

1993	Born in Amsterdam, the Netherlands.

Education

2019–2024 PhD Aerospace Engineering
Delft University of Technology
Promotors: Prof.dr.ir. M. Mulder
Dr.ir. C. Borst

2015–2019 MSc Aerospace Engineering
Delft University of Technology
Thesis: Supplementing Haptic Feedback through the Visual Display of Flight Envelope Boundaries

2011–2015 BSc Aerospace Engineering
Delft University of Technology

Experience

2007–present Developer
FlightGear Flight Simulator

2017 Intern
Royal NLR – Netherlands Aerospace Centre
Focus: Take-off performance monitoring system

Award

2021 Stanley Nelson Roscoe Best Student Paper (for Chapter 3)
International Symposium on Aviation Psychology

List of publications

Included in this thesis.

Awarded the *Stanley N. Roscoe Student Paper Award*.

Journal publications

3. C.A. Badea, **G. de Rooij**, C. Borst, and M. Mulder, *Gamification in Automated Air Traffic Control: Increasing Vigilance Using Fictional Aircraft*. European Journal of Transport and Infrastructure Research, 2025. (accepted)

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

1. **G. de Rooij**, D. Van Baelen, C. Borst, M.M. van Paassen, and M. Mulder, *Supplementing Haptic Feedback in Flight Envelope Protection Through Visual Display Indications*. Journal of Aerospace Information Systems 20(6), pp. 351-367, 2023. doi:10.2514/1.I011191.

Conference publications

7. **G. de Rooij**, A. Stienstra, A.B. Tisza, C. Borst, M.M. van Paassen, and M. Mulder, *Contributing Factors to Flight-Centric Complexity in En-Route Air Traffic Control*. ATM Seminar, Savannah, GA, USA, 2023.

6. **G. de Rooij**, C. Borst, M.M. van Paassen, and M. Mulder, *Preventing Scenario Recognition in Human-in-the-Loop Air Traffic Control Research*. 22nd International Symposium on Aviation Psychology, Rochester, NY, USA, 2023.

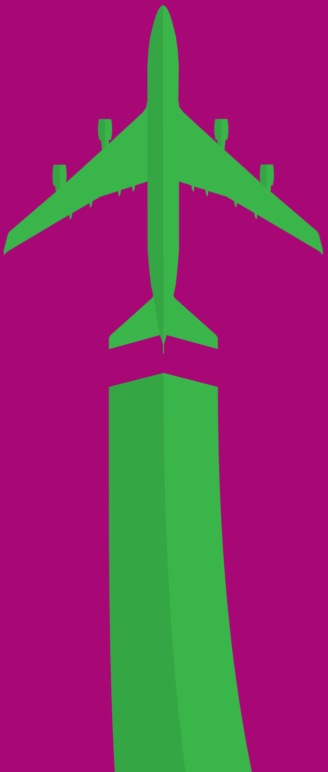
5. M.M. van Paassen, C. Borst, M. Mulder, **G. de Rooij**, F. Dijkstra and A.B. Tisza, *On the Appropriate Participant Expertise for Display Evaluation Studies*. 22nd International Symposium on Aviation Psychology, Rochester, NY, USA, 2023.

4. C.A. Badea, **G. de Rooij**, C. Borst, and M. Mulder, *Gamification for Increased Vigilance and Skill Retention in Highly Automated Air Traffic Control*. SESAR Innovation Days, Budapest, Hungary, 2022.

3. **G. de Rooij**, A.B. Tisza, C. Borst, M.M. van Paassen, and M. Mulder, *Towards Human-Automation Teamwork in Shared En-Route Air Traffic Control: Task Analysis*. 3rd IEEE International Conference on Human-Machine Systems, Orlando, FL, USA, 2022. doi:10.1109/ichms56717.2022.9980715.

2. **G. de Rooij**, C. Borst, M.M. van Paassen, and M. Mulder, *Flight Allocation in Shared Human-Automation En-Route Air Traffic Control*. 21st International Symposium on Aviation Psychology, Online, 2021.

1. **G. de Rooij**, D. Van Baelen, C. Borst, M.M. van Paassen, and M. Mulder, *Supplementing Haptic Feedback Through the Visual Display of Flight Envelope Boundaries*, AIAA Scitech 2020 Forum, Orlando, FL, USA, 2020. doi:10.2514/6.2020-0373.



Air traffic control is transitioning towards a system where human controllers are increasingly supported by high(er) levels of automation. Full autonomy is not within reach in the short term and intermediate levels elicit human-automation issues as human involvement decreases. Applying a high level of automation to only a subset of (low-complexity) flights is hypothesized to address these issues.

Through empirical research, this thesis addresses one of the key challenges for such a system: how to determine which flights should be allocated to either the human or the automation.

The results show the promising effects and general feasibility of applying higher levels of automation to a constrained environment (i.e., a subset of flights). Interacting flights are best allocated to a single agent.