Faculty of Aerospace Engineering

Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control

Increasing Acceptance and Decreasing User Frustration

Maaike Jans

May 23, 2017

TU Delft Delft University of Technology

Challenge the future

Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control

Increasing Acceptance and Decreasing User Frustration

Master of Science Thesis

For obtaining the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

Maaike Jans

May 23, 2017

Faculty of Aerospace Engineering · Delft University of Technology



Delft University of Technology

Copyright © Maaike Jans
All rights reserved.

Delft University Of Technology
Department Of
Control and Simulation

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled **"Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control"** by **Maaike Jans** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: May 23, 2017

Readers:

prof.dr.ir. M. Mulder

dr.ir. M. M. van Paassen

dr.ir. C. Borst

dr.ir. J. de Winter

Acronyms

AA	Adaptive Automation
AI	Artificial Intelligence
ATC	Air Traffic Control
ATCo	Air Traffic Controller
CARS	Controller Acceptance Rating Scale
CPA	Closest Point of Approach
ERATO	En-route Air Traffic Organiser
FPS	Flight Progress Strip
HCA	Human-Centered Automation
ISA	Instantaneous Self Assessment
ITS	Intelligent Tutoring System
LOA	Level of Automation
LoS	Loss of Separation
MTCD	Medium-Term Collision Detection
OOTL	Out-of-the-Loop
PVD	Plan View Display
SA	Situational Awareness
SESAR	Single European Sky ATM Research
SSD	Solution Space Diagram
STCA	Short-Term Collision Alert
URET	User Request Evaluation Tool

Contents

Acronyms v

I Master of Science Thesis Paper 1

II Thesis Book of Appendices 25

A Literature study 27

A-1 Automation transparency . . . 27

A-1-1 Human-centered automation . . . 27

A-1-2 Automation opacity/transparency . . . 28

A-1-3 Automation concepts . . . 29

A-1-4 Automation transparency in Air Traffic Control . . . 33

A-1-5 The transparency of the AA tool of IJtsma (2016) . . . 35

A-2 Effect of transparency on acceptance and frustration . . . 38

A-2-1 Building trust through transparency . . . 38

A-2-2 Decreasing frustration through transparency . . . 39

A-3 Using preview information to achieve transparency . . . 39

B Display adaptations 41

B-1 Workings of the AA tool of IJtsma (2016) . . . 41

B-2 Separation monitor . . . 44

B-3 Increasing transparency . . . 45

B-4 Separation monitor preview functionality . . . 46

B-5 Projected aircraft path functionality . . . 47

C Experiment design **51**
C-1 Experiment outline . . . 51
C-2 Research question and hypotheses . . . 52
C-3 Experiment variables . . . 52
C-3-1 Independent variables . . . 52
C-3-2 Dependent variables . . . 53
C-3-3 Control variables . . . 54
C-4 Experiment matrix . . . 54
C-5 Participants . . . 55
C-6 Participant briefing and training . . . 55

D Checklist for Trust between People and Automation **57**

E Pre-experiment briefing **63**

F Additional experiment results **73**
F-1 Average extra pathlength . . . 73
F-2 Variance of Z-scored ISA ratings . . . 73
F-3 Average maximum distance between aircraft . . . 74
F-4 Number of LOA changes . . . 75
F-5 Total number of commands . . . 76

G Questionnaire results **77**
G-1 NASA-TLX . . . 77
G-2 SASHA . . . 77
G-2-1 Statistical results . . . 78
G-2-2 Comments . . . 78
G-3 Checklist for trust between people and automation . . . 81
G-4 Frustration questionnaire . . . 85

Bibliography **91**

Part I

Master of Science Thesis Paper

Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control

M. Jans, C. Borst, M.M. van Paassen, M. Mulder*

***Abstract*—As air traffic controller workload is a bottleneck on air traffic growth, automation solutions have been proposed. This study investigates the effect of transparency on the acceptance of resolution advisories generated by an adaptive automation tool and the frustration experienced by controllers when using this tool. Two different kinds of transparency were looked at. The first shows the variables the automation uses directly, while the second shows them only indirectly. Both incorporated a preview functionality, which allowed for comparison of the resolution advisory and the controller solution when the automation activated. A human in the loop experiment featuring these different kinds of transparency was performed. The results show that there is no significant difference in acceptance of advisories, workload or frustration ratings obtained from the NASA-TLX between the different levels of transparency. However, the group using the direct form of transparency received more short-term collision alerts and the time that the automation was active was higher than for the group using the indirect transparency. No significant difference in controller trust in the automation was found. In conclusion, there does not appear to be an influence of transparency on controller acceptance of resolution advisories, controller workload, controller frustration or controller trust in the automation. However, the experimenter noticed that most participants in the experiment did not use the information from the preview functionality to compare their solution to the resolution advisory. As they did not use the automation transparency to its full potential, further research on automation transparency is recommended in order to either confirm or dismiss the findings of this study.**

***Index Terms*—Automation transparency, separation monitor, automation acceptance, frustration, trust, air traffic control.**

I. Introduction

Despite a reduction in growth of air traffic recently, predictions show that air traffic in Europe will likely grow by a factor of 1.5 by the year 2035 [1]. As air traffic grows, Air Traffic Controller (ATCo) workload increases. Currently, the ATCo workload is a bottleneck on the growth of air traffic [2]. In order to better manage ATCo workload, automation solutions have been proposed [3, 4, 5]. However, standard automation assists the ATCo with a task under both high and low workload conditions. This may lead to situations in which workload is too low, which can cause vigilance problems [6].

Adaptive Automation (AA) is a type of automation in which decisions on the use, disuse and level of support are shared between the human operator and the automation [7]. In designing this AA, decisions have to be made on when and how the automation should support the ATCo. Most triggering mechanisms considered for AA attempt to detect high workload conditions and react accordingly. High workload conditions can be identified by the occurrence of a certain event, a real-time assessment of operator workload, psychophysiological measurements, or a performance model [8, 9, 10, 11]. However, this results in the AA helping the operator when it might already be too late. Predictive measures can be used to predict when the workload will increase beyond a certain threshold, enabling the automation to react proactively.

IJtsma proposed such a predictive measure in his research, namely the number of 'bad' decisions the ATCo makes [12]. Studies have shown that a significant part of air traffic complexity is the result of suboptimal control decisions [3]. The AA designed by IJtsma intervenes when the ATCo tries to implement a suboptimal solution to a conflict. It was hypothesized that the number of 'bad' decisions is both an indicator and a predictor for high workload, as high workload might cause the ATCo to make worse decisions and those 'bad' decisions cause an even higher workload. This hypothesis was tested in an experiment in which the automation activated after 1, 2 or 3 'bad' decisions and a scenario in which it did not activate at all. When the automation activated, it gave resolution advisories that were consistent with the criteria by which the automation detected 'bad' decisions. It was expected that the automation balances workload, reducing workload and airspace complexity at times when workload is high [12].

The results of the experiment showed that workload remained constant over all experiment conditions. The NASA-TLX shows that workload originating from sources like mental demand and effort are reduced slightly under the stricter automation conditions. However, the decrease in workload from these sources is compensated for by an increase in workload originating from frustration. It was furthermore noted that ATCos would 'fight' against the automation. The ATCos reported that they felt the automation interfered with their own plans. In these situations the automation proposed a better solution, but the ATCo was executing a plan they had made that went several additional steps into the future. As the automation's solution was not in line with this plan, the ATCos felt that their own plan was more efficient on the long run [12].

Past studies have shown that humans sometimes question the accuracy and effectiveness of the automation's actions or advice due to the their difficulties understanding the state of the automation and the reasoning behind its behaviors [13]. If the automation's actions are comprehensible to the operator and the reasoning behind the decision is transparent, the operators may view the automation more favorably [14]. Increasing system transparency might prevent the operator perceiving the automation as trying to inhibit or prevent the realization of their goal, especially when this prevents a conflict between

*All authors are with the Faculty of Aerospace Engineering, Delft University of Technology, Netherlands. E-mail: m.g.m.jans@student.tudelft.nl

the operator and the automation.

This study investigates the effect of transparency on the acceptance of the resolution advisories generated by the AA tool developed by IJtsma, as well as the effect on the frustration level experienced by the ATCos. In this context, two different levels of transparency will be considered. One level of transparency will show information on why the automation activates, using a display showing the exact same variables the automation uses, thus showing the reasoning of the automation. The other level of transparency shows the same information as the first, but in a different format. The second level of transparency does therefore not show the reasoning of the automation. This study explores the effects of showing the reasoning of automation in the context of automation transparency. It is hypothesized that the showing of the reasoning of the automation improves ATCo acceptance of the tool and results in a larger reduction in ATCo frustration.

II. Theoretical background

This section provides a theoretical background for the study presented in this paper. The first part will look into the concept of frustration and the role automation transparency can play in eliminating or reducing frustration. The second part gives an overview of the concept of automation transparency, to gain insight into the concept that will be studied in this paper.

A. Frustration

Frustration arises when there is a hindering condition that hampers or prevents the attainment of a goal [15]. However, according to Maslow, a distinction can be made between deprivation and threat [16]. There is a difference between a deprivation that is not important to the individual, something that is easily replaced or which has few serious consequences, and a deprivation that is a danger to the person, their goals in life, their defensive system, their self-confidence or their sense of security. The effect of frustration can also be found as a result of other types of threat, conflict, rejection, etc. [16].

Responses to frustration by humans can be either adaptive or maladaptive. Adaptive responses are constructive and are aimed at solving the problem that is preventing the goal attainment. Maladaptive responses consist of the absence of constructive problem-solving and usually create additional problems, thereby making the frustrating experience worse. Factors that influence the level of frustration experienced by users of computer systems include the level of dedication to the goal, the severity of the disruption and the strength of will to obtain the goal. A major cause of user frustration with computers can be attributed to bad or confusing interface design [15].

Human frustration with automation is not new; as long as there has been automation, there have been humans frustrated with some factor of this automation. Lin, Yenn and Yang performed an experiment in which the participants had to perform a nuclear plant shutdown task in either a manual or an automated mode [17]. The results of this experiment show that significantly higher levels of frustration occurred in participants under the automatic control mode. In the open questions, a participant stated that they felt out of the loop during the shutdown task in automated mode and that this may have been the cause of the frustration of this particular participant. Kaber, Riley, Tan and Endsley state that an increase of task requirements beyond the human operator's cognitive capabilities can lead to a feeling of frustration and defeat, as well as leading to a loss of self-confidence in the ability to perform the task [18].

Increasing system transparency might prevent the operator perceiving the automation as trying to inhibit or prevent the realization of their goal, especially when this prevents a conflict between the operator and the automation. Furthermore, increased transparency may encourage adaptive responses whenever frustration does occur.

B. Automation opacity/transparency

Automation transparency can be defined as the level of detail with which the automation communicates the reasoning behind its advice on, or solution to, a problem to the human operator [3]. Opacity is the opposite of transparency. High opacity can cause the operator to be unable to develop a good mental model of the automation. However, some level of opacity is always needed in order to avoid overloading the operator with information and data [19].

Höök has identified three distinct types of transparency, namely domain transparency, internal transparency and embedding transparency [20]. Automation that features domain transparency allows the operator to see through the automation and look at the domain behind it. Automation that features internal transparency allows the operator to see through the automation's interface and observe the internal workings of the automation. Embedding transparency refers to transparency of the entire environment in which the automation will be used.

A way to reduce automation opacity is to make the automation 'think' like the human operator about activities and let the automation provide continuous and detailed feedback on its activities. This means that the structure of the automation and its process knowledge should be compatible with that of the human operator [21].

Previous research has shown that an explanation of the advice of an art recommendation system increased the acceptance of the recommendations. However, an increase in trust in the system was not found [22]. Furthermore, Al-Jabri and Roztocki have found a direct significant effect of transparency on perceived usefulness and ease of use of enterprise resource planning systems [23].

III. Displays supporting transparency

A. Separation monitor

The separation monitor is an air traffic control system developed by NATS. It consists of a trajectory prediction module and a conflict detector. The trajectory prediction module calculates a trajectory for each aircraft and inputs aircraft detected position data. The conflict detector uses this information to detect future circumstances under which pairs of aircraft violate the separation minimum and shows this information on a display device [24].

Figure 5 shows the radar screen the ATCO sees, called the Plan View Display (PVD), and the corresponding separation monitor display [24].

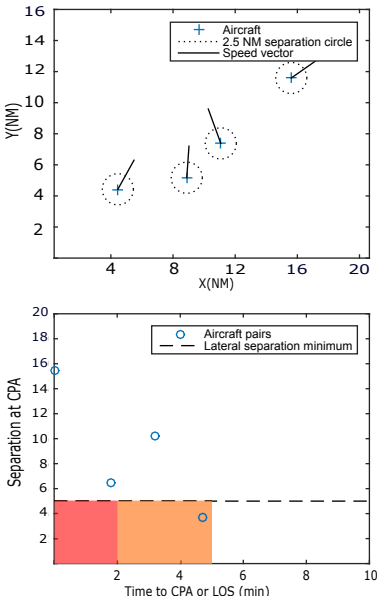


Figure 1: The PVD and the corresponding separation monitor [12].

In figure 5, each circle represents a pair of aircraft. When an aircraft pair appears below the dashed line, it will violate the separation minimum at some time in the future, which is indicated on the x-axis, with the minimum distance between the aircraft indicated on the y-axis. When the aircraft pair will violate the separation minimum at some point in the future, the distance and time indicated are the distance at closest point of approach (CPA) and time to loss of separation (LoS). When the aircraft pair will not violate the separation minimum at some point in the future, the distance and time indicated are the distance at CPA and time to CPA. The red area indicates the occurrence of a short-term collision alert (STCA) for a particular aircraft pair and the orange area indicates the look-ahead time of the adaptive automation developed by IJtsma [12]. The red and orange areas combined constitute the critical area of the separation monitor.

B. The adaptive automation tool of IJtsma

Figure 2 shows the flow diagram for the software of the AA tool. At every point in time, the automation maintains and updates a separation monitor. The automation uses the separation monitor data to determine which aircraft pairs will result in a conflict [12].

Whenever the controller selects an aircraft, if this aircraft has a future conflict with another aircraft, the automation will

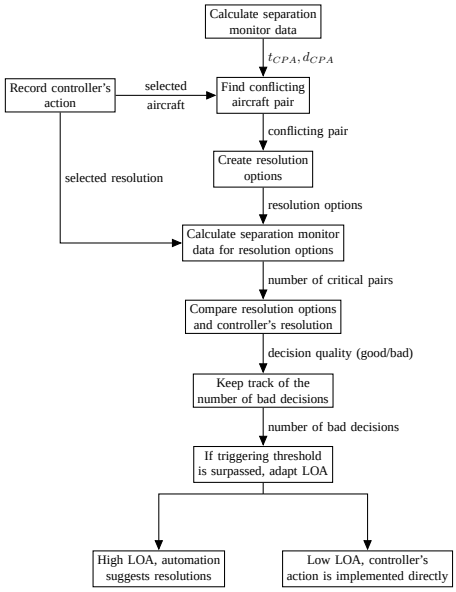
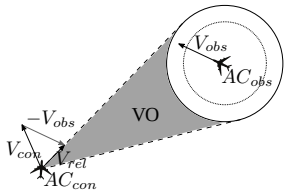
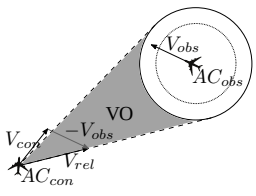


Figure 2: Flow diagram for the AA tool of IJtsma [12]

create a resolution advisory for this aircraft pair. The automation considers four possible solutions, namely a heading change to the left and right for both of the aircraft in the conflict. Figure 3 shows the construction of the calculation of these solutions for one of the aircraft in a pair.



(a) First resolution option



(b) Second resolution option

Figure 3: Construction of resolution options [12]

In Figure 3, a velocity obstacle (VO), which is the set of all relative velocities that will result in a LoS, can be identified [25]. In order to accommodate for controller inaccuracies,

as the controller will most likely not be able to solve a conflict with zero margin, an additional margin (outer circle) was used on top of the minimum required separation (inner circle). To calculate the resolution options, the relative speed vector of the two aircraft is positioned on the boundary of the VO. The resulting vector represents a conflict-free trajectory with a minimum heading deviation. [12]. In Figure 3a, AC_{con} is directed in front of AC_{obs}, while in Figure 3b, AC_{con} is directed in behind AC_{obs}.

As different types of conflicts can have different optimal solutions, IJtsma divided conflicts over three different categories, which can also be seen in Figure 4 [12].

- Same track conflicts (less than 45 or more than 315 degrees heading difference)
- Reciprocal tracks conflicts (more than 135 and less than 225 degrees heading difference)
- Crossing tracks conflicts (heading difference between 45 and 135 or 225 and 315 degrees)

Kirwan et al. identified that for crossing conflicts ATCos generally steer the slower aircraft behind the faster aircraft. However, for same track conflicts, the faster aircraft is often vectored direct to route, in front of the slower aircraft. A standard solution for reciprocal track conflicts was not found [28]. The automation uses this information when determining the best solution to a conflict [12].

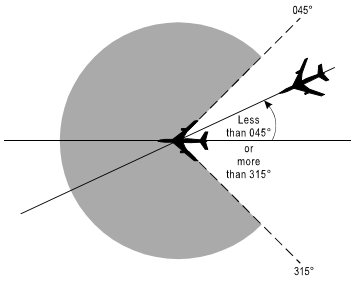
The automation constructs a cost function for each of the four possible solutions. This cost function consists of five factors [12]:

1) The change in the number of aircraft in the critical area of the separation monitor (ΔN)
2) Time to Closest Point of Approach (CPA) (t_{CPA})
3) Heading deviation ($\Delta\Psi$)
4) Separation at CPA (d_{CPA})
5) Speed (V)

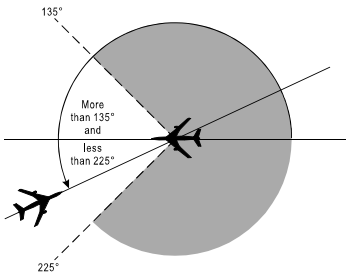
For the different categories of conflicts, different weights are used for the factors in the cost function and these can be found in Table I [12].

TABLE I: Weights for each factor in the cost function in the AA tool of IJtsma [12]

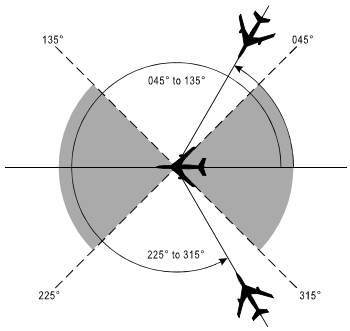
	Factors				
Conflict type	ΔN	t_{CPA}	$\lvert\Delta\Psi\rvert$	d_{CPA}	V
Same track	10	0	0.2	0	-0.02
Crossing	10	0.65	0.02	-0.2	0.01
Reciprocal	10	0.55	0.1	0	0



(a) Same track



(b) Reciprocal tracks



(c) Crossing tracks

Figure 4: Conflict classification as used by ICAO [26]

As the purpose of the automation is to prevent excessive workload by preventing or minimizing the number of induced conflicts, the change in the number of conflicts has a weight that is more than an order of magnitude larger than that of all other factors in the cost function. The best solution to the conflict is found by minimizing this cost function.

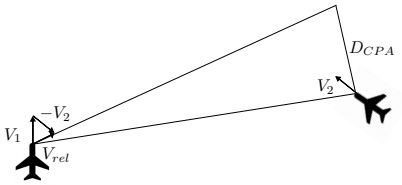
An updated separation monitor is then calculated for both the controller solution and the automation solution and the results are compared. If the controller solution has more induced conflicts than the automation solution, this is recorded as a 'bad' solution. The automation keeps track of the number of 'bad' solutions and will switch to the high level of automation if the triggering threshold is surpassed. When the high level of automation is active, the automation actively provides its solutions to the controller, who can either accept or reject the solution within fifteen seconds. When there is no response from the controller within these fifteen seconds, the solution will automatically be implemented [12].

The transparency of this adaptive automation tool can be increased by showing the separation monitor. As the time to and distance at CPA are used in the calculation of resolution options and the separation monitor itself is used to determine

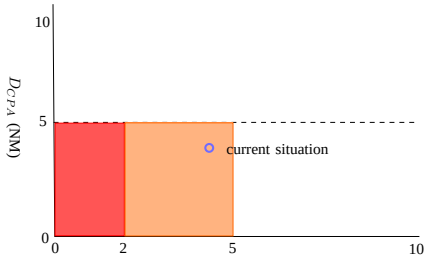
the number of induced conflict, it is the perfect candidate to show the reasoning of the automation. The projected aircraft path function is also capable of showing time to and distance at CPA, but does so by showing projections on the PVD. It therefore does not show the reasoning of the automation directly. The use of the separation monitor in increasing transparency will be discussed in section III-C and the use of the projected aircraft path function in section III-D.

C. Adapted separation monitor

Figure 5a shows a conflict between two aircraft, the distance at CPA between the aircraft and the corresponding separation monitor display [24].



(a) Conflict geometry and distance at CPA

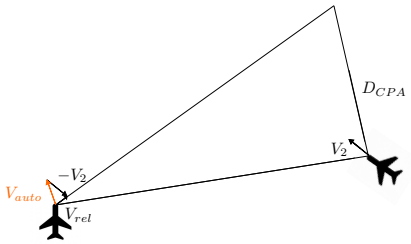


(b) Separation monitor

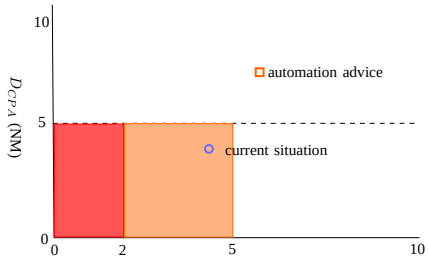
Figure 5: The conflict geometry and the corresponding separation monitor

In Figure 5b, the circle represent the pair of aircraft. In order to help the operator understand the reason for the automation's advice, a preview function has been implemented in the separation monitor. This function shows a preview of the separation monitor when the resolution advisory would be accepted and another preview of the solution the controller suggests. In this way, the controller can see the result of their actions and the results of the resolution advisory and compare them. The preview functionality of the separation monitor can be seen in Figures 6 and 7.

Figure 6a shows the heading the automation proposed for the aircraft and the resulting distance at CPA. In Figure 6b, the orange square indicates the time and distance to CPA if the heading change proposed by the automation would be implemented. In this case, it can be seen that the distance at CPA is larger than 5 NM and the conflict would be solved with this heading change.

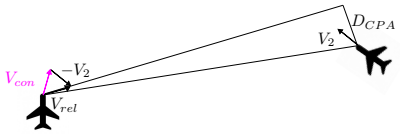


(a) Conflict geometry and distance at CPA

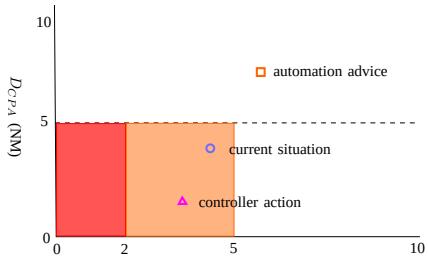


(b) Separation monitor

Figure 6: The conflict geometry and the corresponding separation monitor for the resolution advisory



(a) Conflict geometry and distance at CPA



(b) Separation monitor

Figure 7: The conflict geometry and the corresponding separation monitor for the controller solution

Figure 7a shows the heading the controller proposed which activated the automation and the resulting distance at CPA. In Figure 7b, the magenta triangle indicates the time and distance to CPA if the heading change the controller proposed would be implemented. In this case, it can be seen that the

solution proposed by the controller would not have solved the conflict and that this is the reason the automation activated and suggested another solution.

During the activation of the automation, all aircraft pairs on the separation monitor will be interactive. This means that clicking on an aircraft pair in the separation monitor highlights both aircraft on the PVD. This helps the controller understand what the automation is trying to achieve with the resolution advisory. Furthermore, this functionality should help create visual momentum for the controllers, as they have to extract and integrate information across different displays and this creates a mental link between the displays. The choice has been made to use triangles and squares to indicate the different sets of data in the separation monitor, as they are easier to distinguish than circles of different colors. As the PVD contains colors to represent the controller solution and the automation solution, namely magenta and orange, these same colors will be used for the separation monitor previews. This will create extra visual momentum [27].

As the separation monitor directly shows the variables the automation uses to judge the controller's decisions and create resolution advisories, it can be said that it provides internal transparency [20]. The separation monitor is a separate display, so it is shown on a second screen. As the separation monitor directly shows the reasoning of the automation, it is expected that it will help controllers understand the automation and therefore experience less frustration. It should furthermore benefit acceptance and performance, as controllers should not have the feeling of having to 'fight' the automation and will trust the automation more, as they understand it.

D. Projected aircraft path function

The projected aircraft path function shows projected future aircraft positions on the ATCo radar screen. The projections assume a linear flight path with constant speed and are available for the current situation. When the automation is active, projections will also be available for the situation when the resolution advisory would be implemented and the situation when the operator's solution would be implemented. An example of these projected future aircraft positions can be seen in Figure 8. Projected positions are available for the same amount of time as available on the x-axis of the separation monitor, which is 10 minutes into the future.

The controller uses his mouse to drag the slider at the left up and down to vary the look-ahead time. The controller can drag the slider for any amount of time they wish, during which time the projections will show on the screen. This projected aircraft path function will always be available to the operator for the current traffic situation. When the automation becomes active, the projected aircraft path function will also be available for the automation solution and proposed operator solutions. The projected aircraft path function will use the same color scheme as the separation monitor.

Contrary to the separation monitor, the projected aircraft path function does not show the variables that the automation uses directly. It instead shows them indirectly, through the projections at a certain time in the future. It can therefore be said that the projected aircraft path function provides domain transparency [20]. As the projected aircraft path function does not show the reasoning of the automation, but does give insight into the variables used by the automation, it is expected that it will help controllers understand the automation slightly and therefore experience less frustration compared to when no transparency is present. It should furthermore benefit acceptance and performance, as controllers should not have the feeling of having to 'fight' the automation. However, it is expected that these effects will not be as large as when controllers can directly observe the reasoning of the automation.

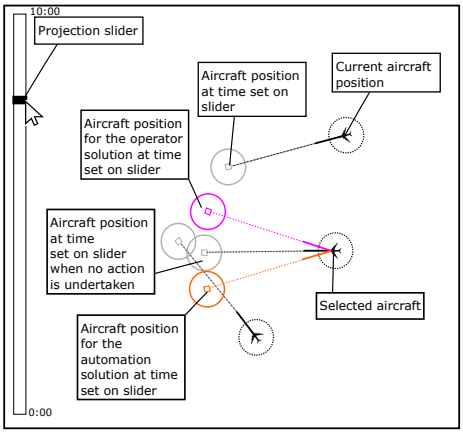


Figure 8: Schematic of the projected aircraft path function

IV. Experiment design

A. Goal

A human-in-the-loop experiment was conducted to determine the effect of the two transparency designs on controller workload, frustration, automation advisory acceptance and safety.

B. Participants and Instructions

Sixteen participants (3 female, 13 male) were selected, consisting of students and staff members of the Faculty of Aerospace Engineering. All subjects had some experience with directing air traffic in an ATC simulation, either through participating in courses or earlier experiments related to ATC.

Participants received a pre-experiment briefing a few days before the experiment in which they were instructed to maintain a separation of at least 5 NM between all aircraft and to clear all aircraft to their exit waypoint, preferably as efficiently as they can. In this briefing, information on the general workings of the simulator and either the separation monitor of the projected aircraft path function was given as well. The adaptive automation tool was introduced as a support tool that will support the participant if and when the automation

itself thinks it is needed. Further information on the triggering mechanism and thresholds was not given.

Some information from this pre-experiment briefing was repeated in a short briefing given before the start of the experiment. Additionally, the controller 'best practices' as identified by Kirwan et al. were introduced to the participant [28]. A paper summary of these best practices was given to the participant to use as a reference during the experiment. During this briefing, the separation monitor or projected aircraft path function was also discussed, to make sure the participant understands the display. Information on the automation was also given to the participants in this briefing, stating that whenever they received an advisory, they should first accept or reject the advisory before they could vector the aircraft themselves.

C. Independent variables

The experiment has two independent variables, namely the AA triggering threshold and the level of transparency. In the experiment of IJtsma, the triggering threshold had four levels, namely a manual control condition and triggering of the AA after 1, 2 or 3 bad operator decisions [12]. These levels were present in this experiment as well. The different levels are:

- Condition AA1: a low threshold condition (triggering after each bad decision), resulting in early automation intervention.
- Condition AA2: a medium threshold condition (triggering after two bad decisions), resulting in intermediate automation intervention.
- Condition AA3: a high threshold condition (triggering after three bad decisions), resulting in late automation intervention.
- Condition MAN: manual control without automation support.

This will allow comparison of the effect of the projected aircraft path function and the separation monitor for the different automation conditions. This variable will be a within-participants factor in this experiment.

The level of transparency has two different levels, namely the projected aircraft path function (PAP) and the separation monitor (Sepmon). This will be a between-participants factor in the experiment. As IJtsma performed his experiment using a baseline level of transparency, namely none, it is not needed to include it in this experiment [12].

D. Control variables

The following control variables will be kept constant:

- The automation will stay active for 30 seconds each time it is triggered. When the automation switches off while an advisory is still active, that advisory remains active until it either expires or is accepted or rejected by the participant.
- Resolution advisories expire after 15 seconds.
- The additional margin that is used in constructing the resolution advisories is set at 2 NM.
- The traffic will be the same for each experiment condition, because this allows comparing the conditions without confound. To prevent the subjects from recognizing conflict geometries from earlier experiment runs, the airspace will be rotated (90, 180 or 270 degrees) between experiment runs. The traffic scenarios used will be the same traffic scenarios used in the experiment of IJtsma, as this allows comparison between the results from this experiment and the experiment of IJtsma without confound.
- The basic separation monitor without preview functionality or interactive capabilities, or the basic projected aircraft path functionality will always be available to the controller, depending on the level of the independent variable transparency.
- The time scale on the x-axis of the separation monitor and the maximum time on the projected aircraft path slider are set to 10 minutes.

E. Dependent measures

The following dependent measures have been recorded:

- Interface
  - The total amount of time the projected aircraft path function is used.
  - The total number of times the projected aircraft path function is clicked on.
  - The total number of times aircraft pairs are clicked on in the separation monitor under active automation.
- Automation acceptance
  - CARS ratings will be asked for after each run [29].
- Subjective workload
  - ISA ratings of workload will be asked for once every minute [30].
  - NASA-TLX ratings will be asked for at the end of each experiment run [31].
- Safety
  - The number and duration of STCAs
  - The number of duration of LoSs
  - The number of mid-air collisions (the number of times during a run when separation decreases below 0.5 NM)
- Automation
  - Automation operative time, which is the percentage of time that automation support is active.
  - The number of advisories, which can be subdivided into accepted, rejected and expired advisories.
- Airspace complexity
  - The relative no-go-area of the solution space can be used as a predictor of workload. This metric indicates how many vector commands in the available control space (defined as all possible vector commands within 360 degrees from the current heading) will result in a conflict with other aircraft. The metric considers the maximum no-go area of all aircraft in the airspace [32].

- Situation awareness
  - An adaptation of the SASHA questionnaire will be completed after each run [33].
- Trust
  - The Checklist for Trust between People and Automation developed by Jian et al. will be completed after each experiment run [34].
- Frustration
  - A frustration questionnaire will be completed by participants after the experiment.

Previous research has indicated that people might unknowingly act 'politely' to a computer when the same computer asks for feedback on itself [35]. Therefore, the NASA-TLX ratings will be completed on a paper form, and the CARS ratings, the SASHA questionnaire and the checklist for trust will be completed on a different computer. It must also be noted that the NASA-TLX ratings include a measure for frustration and that this and the frustration questionnaire will be the only measures of frustration used in the experiment. As the checklist for trust was meant for native English speakers, a Dutch translation of the checklist was made.

F. Apparatus

The experiment was conducted in the Air Traffic Management Laboratory (ATMLab) of the Control and Simulation Division at the Delft University of Technology. This lab uses LCD screens on which a simulated PVD is shown, a schematic of this PVD can be seen in Figure 9.

In Figure 9, the polygon indicates the border of the airspace sector, with waypoints on the corners. Aircraft move around in this sector. At the top left of the PVD, the simulation time, a performance score and the run number can be seen. When the automation is active, an indication will show beneath this box and another box will be shown at the top right of the screen. This box contains a list of all advisories and buttons to either accept or reject the advisories.

With this simulation, participants could interactively control aircraft using the mouse and keyboard. Heading changes could be given to aircraft by clicking on an aircraft and subsequently clicking a different heading. When the subject pressed the ENTER key, the heading change was sent to the aircraft and implemented immediately (corresponding to a situation in which ADS-B is used to communicate and implement vector commands).

G. Scenarios

The traffic scenarios used in this experiment were the same scenarios as used in the experiment of IJtsma [12]. A schematic of the traffic scenario used can be found in Figure 10.

The traffic consists of a stream of aircraft flying at 180 kts and two streams of aircraft flying at 220 knots crossing the first stream at an angle of 90 degrees. This geometry was chosen because the crossing conflict has the most clearly identified best practice, so controller and automation solutions should be most similar for this type of conflict [12].

H. Experiment procedure

The experiment consisted of a training phase and an experiment phase. There was a break between the training and experiment phases. If the participant indicated a need for more time, another break was held between the experiment runs.

In the training phase, the participants were presented with eight training scenarios of increasing difficulty. In the first two scenarios, the workings of the simulator and automation were explained again and the participant could try out all functions. In the third scenario, participants were told that they could activate the automation in a particular way and that this scenario was meant for them to get more familiar with the changes in the interface upon activation of the automation. From the fourth scenario onward, the participants were encouraged to try and manage the traffic to the best of their ability.

After training runs five to eight, the participants were asked to fill out the NASA-TLX, adapted SASHA questionnaire, CARS and checklist for trust. This ensured that the participants understood all questions on these questionnaires and had a few occasions on which to ask for explanation on aspects they did not understand. After completing the eight training runs, participants were considered sufficiently trained to continue with the measurement phase.

In the measurement phase, participants were presented with four measurement scenarios, one for each experiment condition. All measurement runs used the same scenario, but the traffic streams were rotated by 0, 90, 180 and 270 degrees for the different runs. All measurement runs lasted 15 minutes. Experiment conditions have been balanced among participants using a Latin square design.

I. Hypotheses

The study presented in this paper focuses on the following research question: What is the influence of transparency on the acceptance of and controller frustration with the AA tool developed by IJtsma [12]? The following hypotheses will be tested in the experiment:

- Availability of the separation monitor results in higher operator acceptance of resolution advisories than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in lower operator frustration than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in lower operator workload than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in higher operator trust in the automation than availability of the projected aircraft path functionality.
- Availability of the separation monitor or the projected aircraft path function results in higher operator acceptance of resolution advisories than when no automation transparency is available.
- Availability of the separation monitor or the projected aircraft path function results in lower operator frustration than when no transparency is available.

9

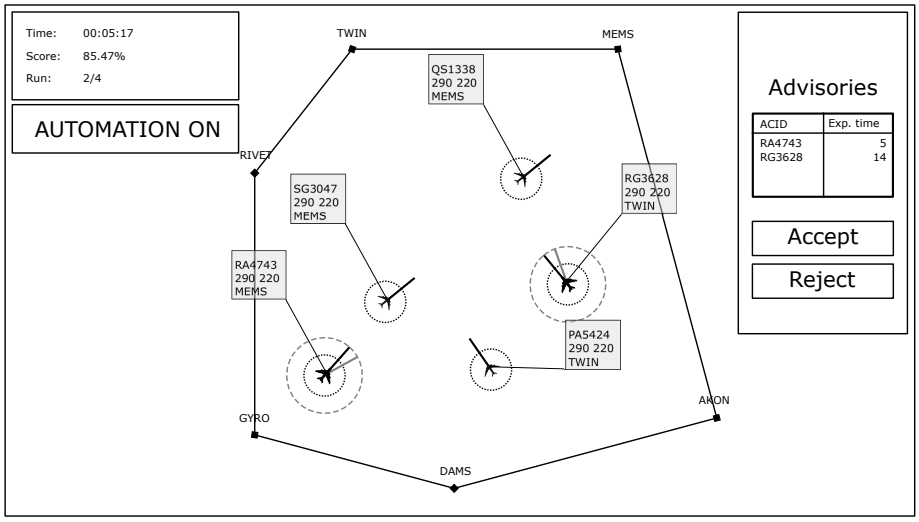


Figure 9: Schematic of the PVD as used in the experiment [12]

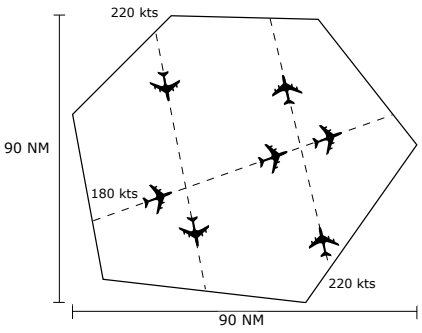


Figure 10: Schematic of the traffic scenario used for the measurement runs (Adapted from [12])

- Availability of the separation monitor or the projected aircraft path function results in lower operator workload than when no transparency is available.

V. Results

This section shows the results of the human in the loop experiment. For comparison to the results from IJtsma (further called the 'baseline' condition as it featured no transparency), 8 participants from his data set (16 participants) have randomly been selected [12]. This has been done to have equal group sizes for all experiment conditions.

A. Interface

Figure 11 shows boxplots of the number of times the separation monitor was clicked on. As the separation monitor could only be clicked on when the automation was active, no clicks were recorded for the manual control condition. It can be seen that the number of separation monitor clicks increases for the stricter automation conditions.

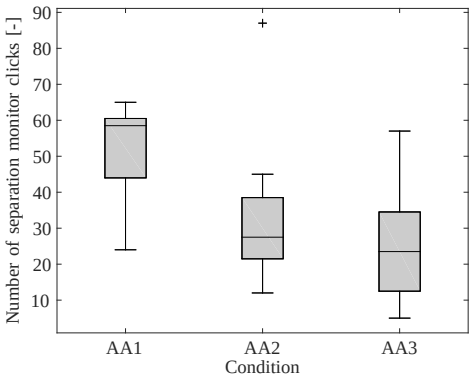


Figure 11: Number of clicks on the separation monitor

Figure 12 shows boxplots of the number of clicks on the projected aircraft path function. It can be seen that the projected aircraft path function was clicked more often than the separation monitor. However, the projected aircraft path function could be used at all times, while the separation monitor could only be clicked on whenever the automation

was active. The numbers can therefore not be compared. From Figure 12, it can also be seen that the projected aircraft path function was clicked on slightly fewer times for the stricter automation conditions.

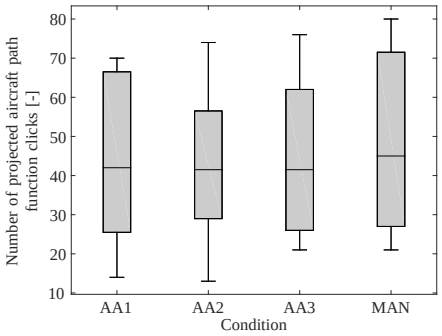


Figure 12: Number of clicks on the projected aircraft path function

Figure 13 shows percentage of time the projected aircraft path function was in use. It can be seen that, despite the function being clicked slightly less often for the stricter automation conditions, the percentage of time the projected aircraft path function was in use is fairly constant over the different automation conditions.

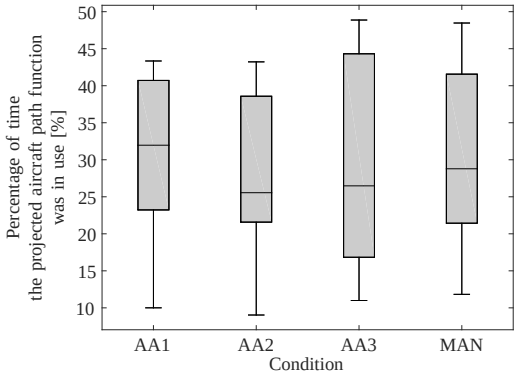


Figure 13: Percentage of time the projected aircraft path function was in use

Despite these results, the experimenter observed that the participants were not using the separation monitor and projected aircraft path function to compare their solution to that of the automation. The separation monitor group appeared to not be able to quickly link an aircraft pair to a symbol on the separation monitor. Whenever the automation was not active, they looked at the separation monitor and identified that there was a conflict coming, but then tended to wait until they received a medium-term collision alert for an aircraft pair before solving the conflict, as they were seemingly unable to identify the aircraft pair belonging to the separation monitor symbol on the PVD. When a heading change was implemented, they would look at the overall situation on the separation monitor to see whether or not their actions were successful in solving the conflict, or had created a different conflict. When the automation was active, they would first accept or reject the advisory without comparing their solution to the resolution advisory. Afterward, they would start clicking on the symbols of upcoming conflicts to identify the aircraft pair belonging to the symbol and try and solve the conflict.

For the projected aircraft path group, there appeared to be no difference in the way the projected aircraft path function was used when the automation was active compared to when it was not active. Participants used the function to probe for conflicts happening in the near future. As soon as a conflict was found, they tried to resolve it, using the function to check the validity of their solution after implementation. When the automation activated, they would first accept or reject the advisory without comparing it to their own solution, to then carry on probing for and solving conflicts.

B. Acceptance

Two different measures of acceptance have been recorded during the experiment, namely the CARS ratings and the percentage of advisories that were accepted. These measures indicate the participants' acceptance of the automation and whether or not their actions are in line with the subjective CARS ratings.

1) CARS: Figure 14 shows the results of the CARS ratings collected after each experiment run. For the Sepmon group, the CARS ratings for condition AA2 appear to be slightly lower than those for conditions AA1 and AA3. Furthermore, the ratings for the manual control condition are very spread out. For the PAP group the CARS ratings for AA1, AA2 and AA3 appear to remain fairly constant, with the ratings for the manual control condition slightly higher. Overall, the CARS ratings for the PAP group appear to have a higher variance than those for the Sepmon group. The CARS ratings for the baseline group appear to be similar to those of the Sepmon and PAP groups for conditions AA1, AA2 and AA3. For the manual control condition, all participants in the baseline group gave a CARS rating of zero, as there was no automation present. Interestingly, participants in the Sepmon and PAP groups gave a nonzero CARS rating for the manual control condition, apparently due to their inability to see the automation and the display as two separate entities.

A Friedman's two-way ANOVA shows that there is no significant difference in CARS ratings between the different automation conditions. When comparing the CARS ratings between the Sepmon, PAP and baseline groups, a Kruskal-Wallis test shows that for conditions AA1, AA2 and AA3, there is no difference between the different conditions. For the manual control condition, all participants in the baseline condition gave the system a rating of zero, resulting in a significant difference between the PAP and baseline and Sepmon and baseline conditions ($H(2)=16.389$; $p<0.05$).

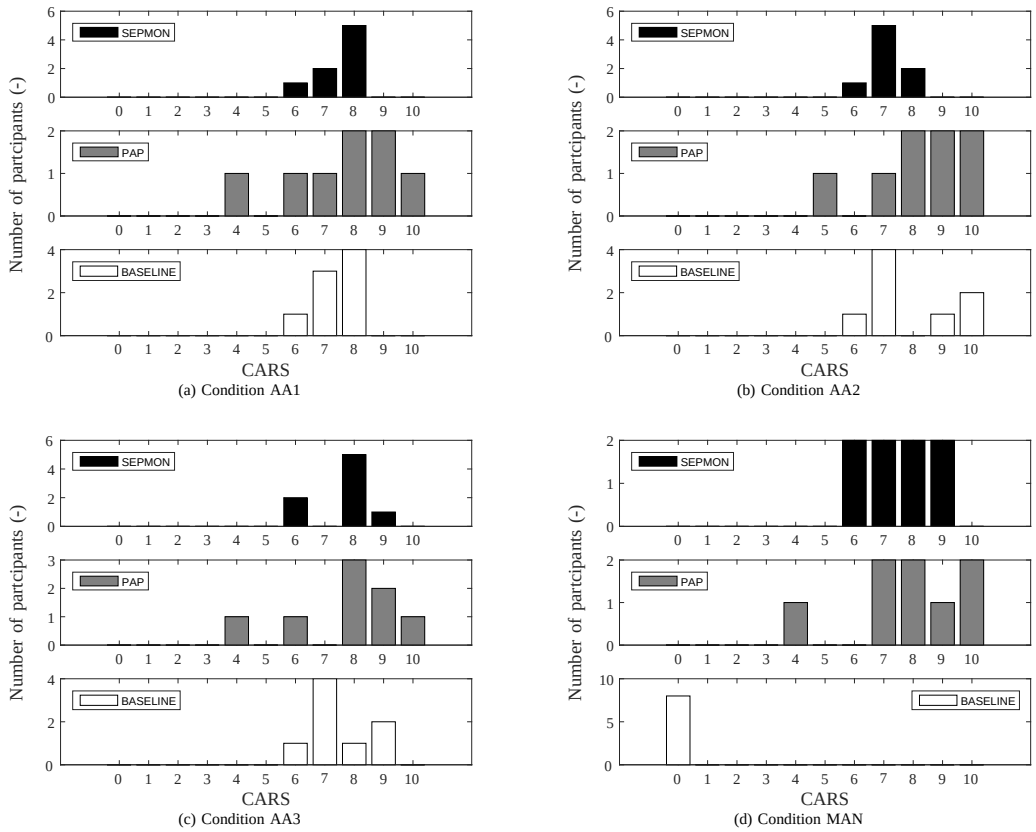


Figure 14: Controller Acceptance Rating Scale (CARS).

2) Advisory acceptance: Figure 15 shows boxplots of the percentage of advisories that have been accepted for the different conditions. It appears that participants in the Sepmon and PAP groups accepted fewer advisories that participants in the baseline group. Acceptance percentages appear to vary most for the Sepmon group and variance in acceptance percentages appears to be lower for the baseline group.

A Friedman's Two-way ANOVA shows that there is no significant difference in acceptance percentages across the different automation conditions. Furthermore, a Kruskal-Wallis test shows that there is no significant difference between the Sepmon, PAP and baseline groups for any of the automation conditions. The results from the acceptance percentages thus match the results of the CARS ratings, showing that there are no significant differences between automation conditions and experiment groups.

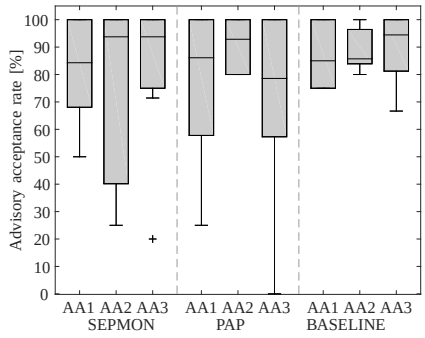


Figure 15: Percentage of advisories accepted

C. Workload

1) ISA ratings: All recorded ISA ratings have been normalized per participant. Figure 16 shows boxplots of the means of these normalized ISA workload ratings. The mean of the ISA ratings was taken as a measure in order to be able to compare the data from this experiment to the baseline condition from the experiment of IJtsma [12]. No apparent trend can be seen in the mean workload ratings for the different automation conditions. However, for the strictest automation condition, use of the PAP function appears to result in less variance in workload ratings between operators, with perhaps a slightly lower mean workload. It can also be seen that the Sepmon group appears to have experienced a lot less variance in mean workload over all automation conditions.

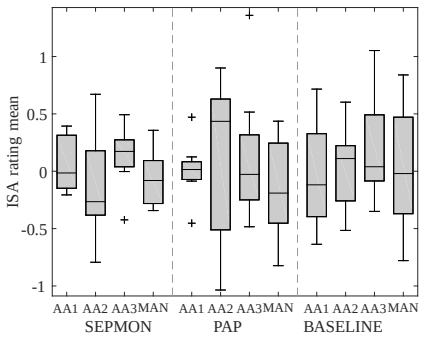


Figure 16: Means of the normalized ISA workload ratings

A repeated measures ANOVA shows that there is no significant difference in mean workload across the different automation conditions. When comparing the mean workload ratings from the Sepmon, PAP and baseline groups, no significant difference is found between the groups.

2) NASA-TLX: Figure 17 shows boxplots of the NASA-TLX workload ratings. It appears that workload ratings decrease for the stricter automation conditions. Furthermore, the workload ratings for the Sepmon group appear to be higher than those for the baseline group, while the workload ratings for the PAP group appear to be lower than those for the baseline group. Higher workload for the Sepmon group might exist because they had to divide their attention between two different computer screens.

A repeated measures ANOVA shows that there is no significant difference in NASA-TLX workload ratings between the different automation conditions. Additionally, no significant difference in workload ratings between the Sepmon, PAP and baseline groups was found.

Figures 18, 19 and 20 show a visualization of the different components of the NASA-TLX averaged over all participants for the Sepmon, PAP and baseline groups respectively. The NASA-TLX consists of 2 parts. In the first part, participants are asked to make pairwise comparisons between all different sources of load. From these comparisons, different weights for each source of load are determined. These weights can

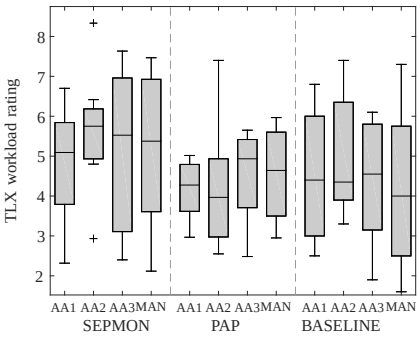


Figure 17: NASA-TLX workload ratings

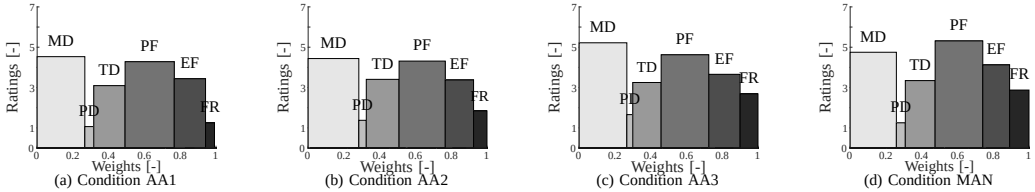
be seen on the x-axis of Figures 18 - 20. In the second part, participants rate the load for each of the sources on a scale of 0 to 10. These are shown on the y-axis of Figures 18 - 20. For the baseline group, physical demand is not shown, because it did not notably contribute to workload (only one subject gave a nonzero weight to physical demand).

For the baseline group, it can be seen that workload from frustration increases in both weight and magnitude for the stricter automation conditions. Interestingly, the opposite appears to be true for the separation monitor and projected aircraft path groups. Here, workload from frustration appears to decrease in both weight and magnitude for the stricter automation conditions. However, a repeated measures ANOVA shows that for the z-scored frustration magnitude ratings, there is no significant difference between the different automation conditions, nor is there a significant difference between the Sepmon, PAP and baseline groups. For the frustration weights, a repeated measures ANOVA shows that there is no significant difference between the different automation conditions, nor is there a significant difference between the Sepmon, PAP and baseline groups.

D. Safety

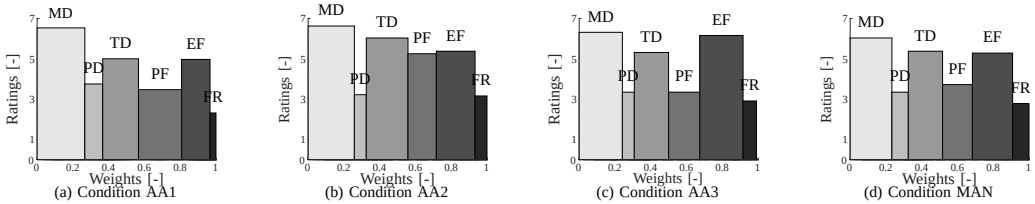
1) Minimum separation: Figure 21 shows boxplots of the minimum separation between aircraft. For the Sepmon and PAP groups, minimum separation appears to decrease slightly for the stricter automation conditions. Overall, differences between the Sepmon, PAP and baseline groups appear to be very small.

A few losses of separation and one mid-air collision occurred in the experiment, as can be seen in Figure 21. For the baseline condition, one LoS occurred, which took place for condition AA1 after an advisory had been rejected. Within a few seconds of it occurring, this LOS was resolved by the controller. For the Sepmon group, the LoSs occurred when the automation gave advisories for conflicts that would result in a LoS within 0.5 minute. This happens when a controller clears an aircraft to it's exit waypoint when close and this results in a conflict with another aircraft close to the same waypoint. Because the automation is unable to accommodate for the turn



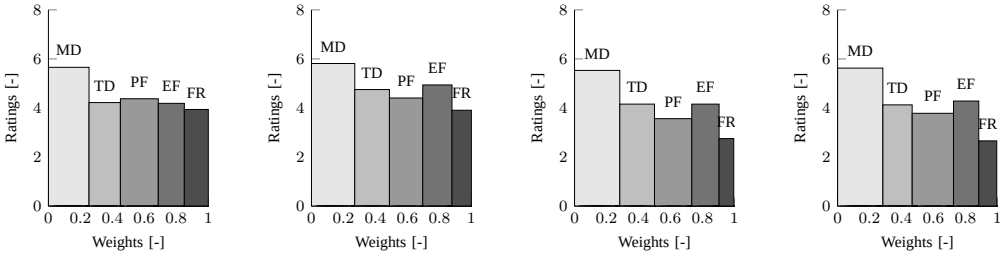
(a) Condition AA1 (b) Condition AA2 (c) Condition AA3 (d) Condition MAN

Figure 18: Visualization of the different components of the NASA-TLX for the Sepmon group. MD = Mental demand, PD= Physical demand, TD = Temporal demand, PF = Performance, EF = Effort, FR = Frustration.



(a) Condition AA1 (b) Condition AA2 (c) Condition AA3 (d) Condition MAN

Figure 19: Visualization of the different components of the NASA-TLX for the PAP group. MD = Mental demand, PD= Physical demand, TD = Temporal demand, PF = Performance, EF = Effort, FR = Frustration.



(a) Condition AA1 (b) Condition AA2 (c) Condition AA3 (d) Condition MAN

Figure 20: Visualization of the different components of the NASA-TLX for the baseline group. MD = Mental demand, PD= Physical demand, TD = Temporal demand, PF = Performance, EF = Effort, FR = Frustration.

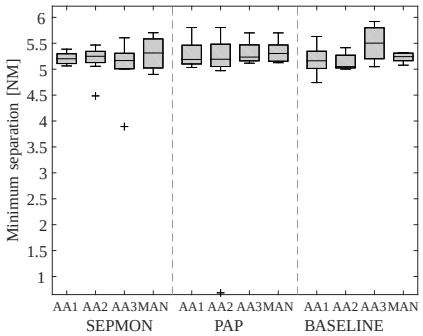


Figure 21: Minimum separation between aircraft

rate of the aircraft, the advisory given by the automation will not solve the conflict. As participants did not notice this and accepted the advisory, a LoS occurred subsequently. These LoSs were quickly resolved by the controller. For the PAP group, the only LoS occurred because the controller cleared an aircraft to it's exit waypoint too early, creating a conflict. The automation did not intervene at this point, because it was set to activate only after 2 'bad' decisions and this was the first. A LoS occurred almost immediately after implementing this heading change and this caused the controller to panic and only be able to solve the conflict after a mid air collision had occurred.

A Friedman's two-way ANOVA shows there is no significant difference in minimum separation for the different automation conditions. Furthermore, a Kruskal-Wallis test shows no significant difference in minimum separation between the

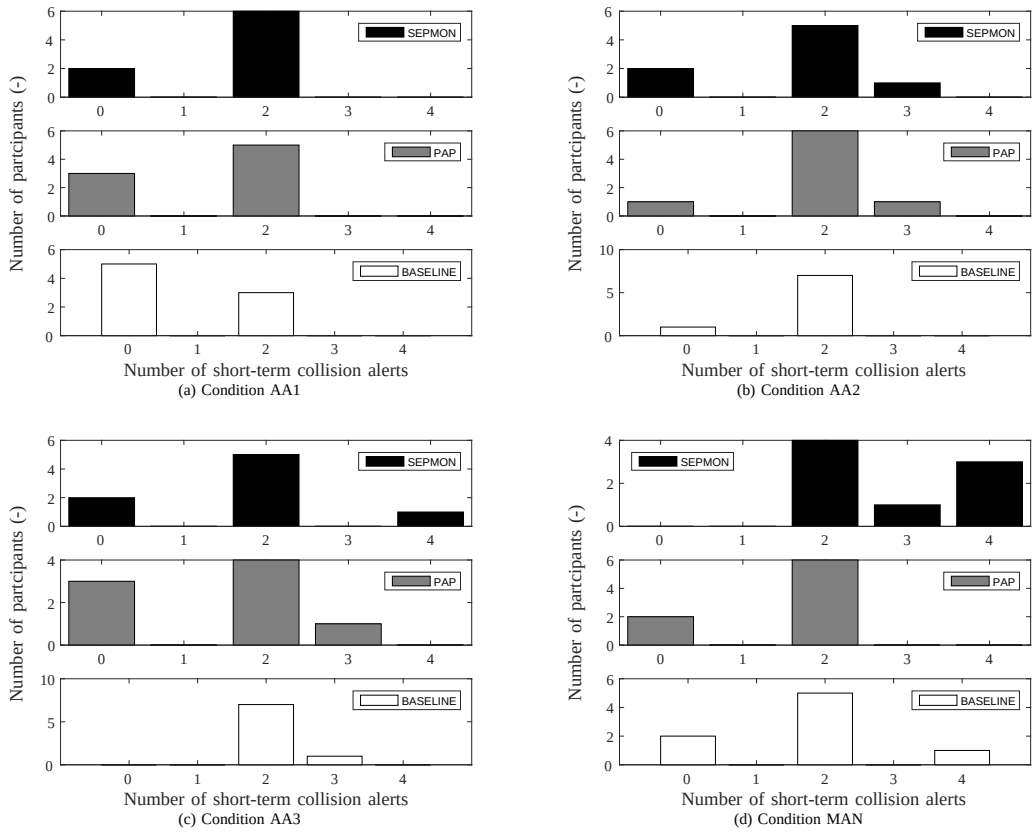


Figure 22: Number of short-term collision alerts

Sepmon, PAP and baseline groups for any of the automation conditions.

2) Number of STCA alerts: Figure 22 shows the number of STCAs for the different automation conditions and experiment groups. It appears that there is a slight decrease in number of STCAs for the stricter automation conditions. Furthermore, it appears that the Sepmon group received slightly more STCAs than the other two groups.

A Friedman's two-way ANOVA shows that there is a significant difference in number of STCAs between the different automation conditions ($\chi^2(3)$=9.214; p $<$0.05). Post-hoc pairwise comparison using the Dunn's procedure reveals that a significant difference in number of STCAs exists between conditions AA1 and MAN (Z=-0.875; p$<$0.05). A Kruskal-Wallis test shows that there is no significant difference in number of STCAs between the Sepmon, PAP and baseline groups for automation conditions AA1, AA2 and AA3. However, a significant difference in number of STCAs between the Sepmon, PAP and baseline group exists for the manual control condition (H(2)=6.504; p$<$0.05). Post-hoc pairwise comparison using the Dunn's procedure reveals that a difference exists between the Sepmon and baseline groups (Z=-6.000; p=0.05) and a difference exists between the Sepmon and PAP groups (Z=-7.312; p$<$0.05).

E. Automation

1) Automation active time: Figure 23 shows boxplots of the percentage of time the automation was active. A clear trend of more automation active time at stricter automation conditions can be seen. Furthermore, it appears that the PAP group had less automation active time for the stricter automation conditions than the Sepmon and baseline groups. The Sepmon group appears to have had more automation active time than the baseline group.

A repeated measures ANOVA confirms that there is a significant difference in automation active time for the different

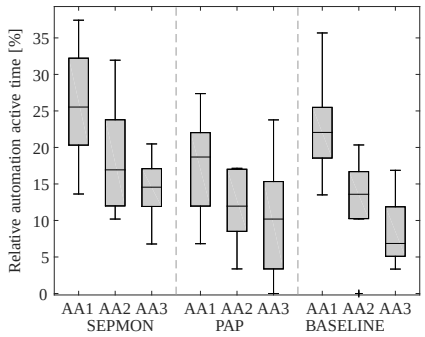


Figure 23: Percentage of time the automation was active

automation conditions ($F(2)=21.385$; $p<0.05$). A significant difference in automation active time is also found between the Sepmon, PAP and baseline groups ($F(2)=4.235$; $p<0.05$). Post-hoc tests using the Bonferroni correction show that a significant difference in automation active time exists between the Sepmon and PAP groups, with a mean of 19.55% active time for the separation monitor group and a mean of 13.19% active time for the projected aircraft path group ($p<0.05$).

2) Number of advisories: Figure 24 shows boxplots for the number of advisories received by participants during a single measurement run. A clear trend of more advisories given for stricter automation conditions is present. Furthermore, it appears that for the stricter automation conditions, the PAP group received less advisories than the Sepmon and baseline groups. However, for the most lenient automation condition, the PAP group appears to have experienced a lot more variance in the number of advisories and have received slightly more advisories on average. Additionally, the Sepmon group appears to have received slightly more advisories than the baseline group.

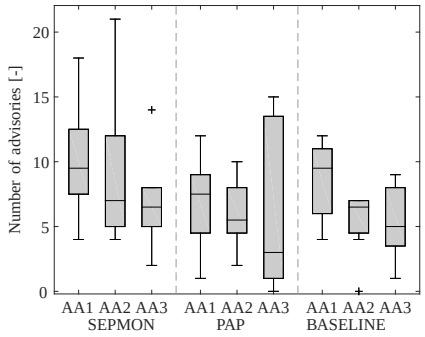


Figure 24: Number of advisories received by participants

A repeated measures ANOVA shows that the difference in the number of advisories given for the different automation conditions is not significant. Furthermore, there is no significant difference in the number of advisories given between the Sepmon, PAP and baseline groups.

F. Airspace complexity

Figure 25 shows boxplots of the mean maximum complexity of the airspace. The mean maxiumum complexity of the airspace is defined as mean of the maximum relative no-go-area of the solution space. This metric indicates how many vector commands in the available control space (defined as all possible vector commands within 360 degrees from the current heading) will result in a conflict with other aircraft. The metric considers mean of the maximum no-go area of all aircraft in the airspace [32]. As the measurement scenarios have a calmer period at both the beginning and the end, the first and last two minutes have not been counted for this metric.

It appears that for the Sepmon and PAP groups, mean maximum airspace complexity increases slightly for the stricter automation conditions. Furthermore, the variance in mean maximum airspace complexity appears to decrease for the stricter automation conditions. Overall, mean maximum airspace complexity appears to be slightly lower for the projected aircraft path group than for the baseline group.

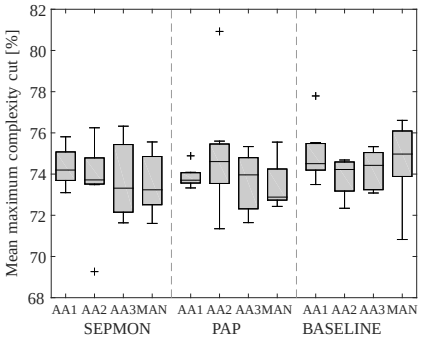


Figure 25: The mean maximum complexity of the airspace

A Friedman's two-way ANOVA shows that there is no significant difference in mean maximum airspace complexity for the different automation conditions. Furthermore, a Kruskal-Wallis test shows that there is no significant difference in mean maximum airspace complexity between the Sepmon, PAP and baseline groups for any of the automation conditions.

G. Questionnaire responses

1) Adapted SASHA questionnaire: Figure 26 shows the results of the adapted SASHA questionnaire. The white boxplots show results for the PAP group, while the gray boxplots show results for the Sepmon group. The SASHA questionnaire was adapted to not incorporate questions on tasks that were not performed in this experiment and incorporate extra questions on the adaptive automation and displays. The questions on this questionnaire are numbered 1 to 12 and are shown below. Questions 1, 2, 9 and 12 relate to high situational awareness,

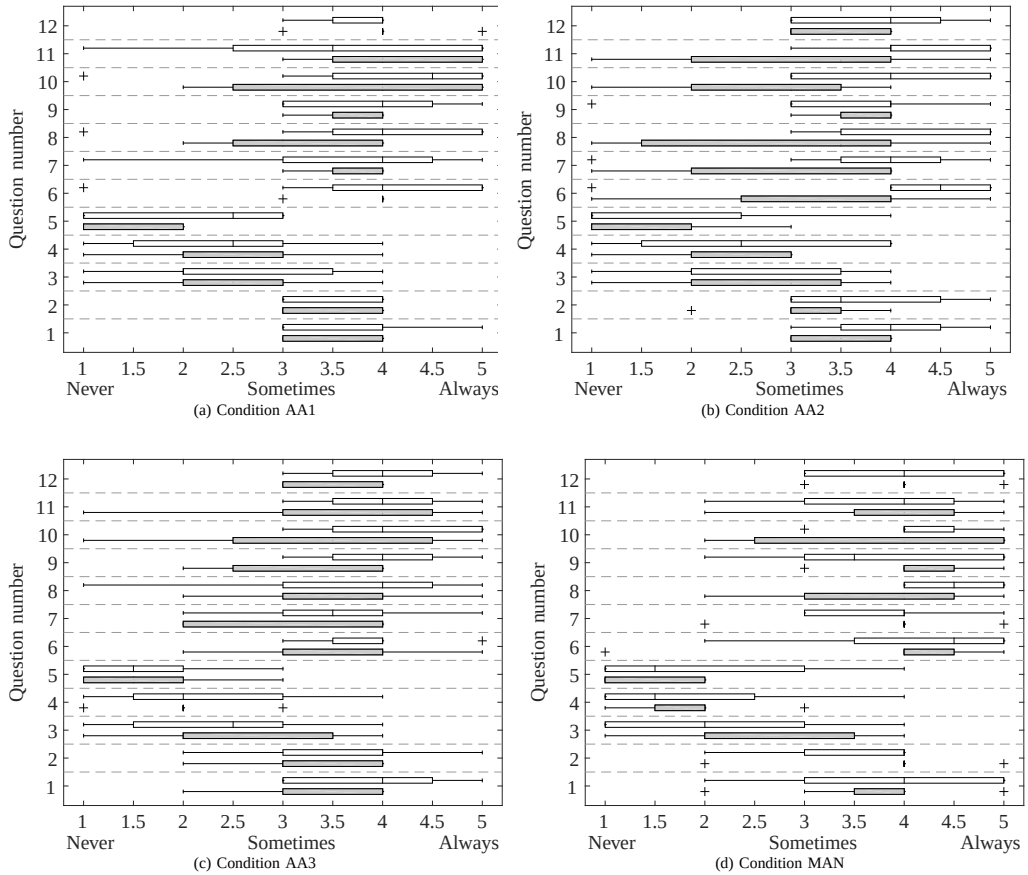


Figure 26: Adapted SASHA questionnaire results for the different automation conditions

while questions 3 to 5 relate to low situational awareness. Questions 6 to 8 relate to the controller opinion on the automation and questions 10 and 11 relate to the controller opinion on the display in relation to the automation.

1) Did you have the feeling that you were ahead of the traffic, able to predict the evolution of the traffic?
2) Did you have the feeling that you were able to plan and organize your work as you wanted?
3) Have you been surprised by a conflict that you were not expecting?
4) Did you have the feeling of starting to focus too much on a single problem and/or area of the sector?
5) Did you forget to vector an aircraft to its exit waypoint?
6) Do you think the adaptive automation provided you with useful information?
7) Did the adaptive automation help you have a better understanding of the situation?
8) Did you feel the adaptive automation provided you support at the right times?
9) Did the interface help you to have a better understanding of the situation?
10) Did the interface help you to understand why the automation triggered?
11) Did the interface help you to understand the intentions of the adaptive automation?
12) Finally, how would you rate your overall situation awareness during this exercise?

In Figure 26, answers to the adapted SASHA questionnaire seem to be fairly similar for the different automation conditions. Furthermore, answers also appear to be similar between the Sepmon and PAP groups, but overall, situation awareness appears to be slightly higher for the PAP group, as this group reported they were able to predict traffic better and reported having more oversight. It can also be seen that the PAP group

appears to find the automation more useful and gave a higher rating on the questions asking if the display helped with their understanding of the automation.

A Friedman's two-way ANOVA shows that there is no significant difference in answers to any of the questions for the different automation conditions between any of the automation conditions. A Kruskal-Wallis test shows that there is no significant difference in answers between the Sepmon and PAP groups for any of the automation conditions for any of the questions, with one exception. When asked the question 'Did the interface help you to understand why the automation triggered?' the PAP group responded with a higher rating than the Sepmon group for the AA2 condition ($H(1)=3.990$; $p < 0.05$).

Comments on the SASHA questionnaire show that overall, participants were satisfied with their situational awareness. Several participants in the PAP group mentioned that the display really helped with their situational awareness. They did not elaborate on why they felt this way. Most participants rated the automation highly, although several of them mentioned that the advisories were not always optimal in their opinion or that they did not have enough time to carefully consider the advisory. Several participants mentioned that it was difficult to relate the symbols on the separation monitor to the aircraft pairs on the PVD.

2) Checklist for trust: The Checklist for Trust between People and Automation consists of twelve statements, of which the first five relate to distrust, while the other seven relate to trust. Answers are given on a scale of 1 to 7, where a 1 stands for not at all and a 7 stands for extremely. The statements on this questionnaire are numbered 1 to 12 and are shown below.

1) The system is deceptive.
2) The system behaves in an underhanded manner.
3) I am suspicious of the system's intent, action, or output.
4) I am wary of the system.
5) The system's action will have a harmful or injurious outcome.
6) I am confident in the system.
7) The system provides security.
8) The system has integrity.
9) The system is dependable.
10) The system is reliable.
11) I can trust the system.
12) I am familiar with the system.

Figure 27 shows the results of the checklist for trust. The gray boxplots show the results for the Sepmon group and the white boxplots show the results for the PAP group.

In Figure 27, answers to the checklist for trust seem to be fairly similar for the different automation conditions. Furthermore, answers also appear to be similar between the Sepmon and PAP groups.

A Friedman's two-way ANOVA shows that there is no significant difference in answers to any of the questions between any of the automation conditions. A Kruskal-Wallis test shows that there is no significant difference in answers between the Sepmon and PAP groups for any of the automation conditions for any of the questions.

It appears that participants trusted the automation very much, regardless of the automation condition or whether they were using the separation monitor or projected aircraft path function. Questions 1 to 5 on the checklist for trust are related to distrust. Participants mostly gave answers between 1 and 3 for these questions, indicating little distrust. Questions 6 to 12 on the checklist for trust are related to trust. Participants mostly gave answers between 5 and 7 for these questions, indicating a high level of trust. The baseline group did not fill out this questionnaire. A correlation analysis shows that there is a negative correlation between the questions related to trust and the questions related to distrust. This was expected, as Jian et al. indicate that trust and distrust are opposites [34].

3) Frustration questionnaire: Thirteen out of sixteen participants from the Sepmon and PAP groups filled out a questionnaire at the end of the experiment, asking them to indicate what frustrated them the most, why this frustrated them and if and how they tried to reduce that frustration. The first three participants did not fill out this questionnaire, as it was only decided to use this questionnaire after these participants had already performed the experiment. Several participants mentioned that the fact that they were forced to respond to the automation every time it activated frustrated them, as they could be very busy at that moment and they were given very little time to respond to the automation. Further reasons for frustration were that the automation did not always advise an optimal solution according to the participants. Because of this, many participants felt that they did not have enough time to consider the solution presented by the automation and either had to accept them without understanding them or reject them and implement their own solution anyway. Most participants stated that they just accepted the advisories without checking the solution first as a way to reduce the frustration they experienced and that they increased the margins they used in their own solutions in order to try and not activate the automation in the first place. Participants did not state whether or not the separation monitor or projected aircraft path function helped them with this change in strategy.

VI. Discussion

From the results, it appears that both the separation monitor and the projected aircraft path function did neither increase acceptance nor decrease frustration. A significant difference in CARS ratings was found between the separation monitor and baseline groups and projected aircraft path and baseline groups for the manual control condition. This is an interesting result, as there was no automation present in the manual control condition. Despite that, the participants in the Sepmon and PAP groups still gave a nonzero CARS rating to the automation. It might be that the participants were unable to see the automation and the displays as separate entities, instead thinking that the displays are part of the automation. Furthermore, during the experiment it was noticed by the experimenter that some of the participants in the Sepmon and PAP groups did not notice that the automation never became active during this scenario and that they still filled out all questionnaires as if the automation had activated at some point during the run.

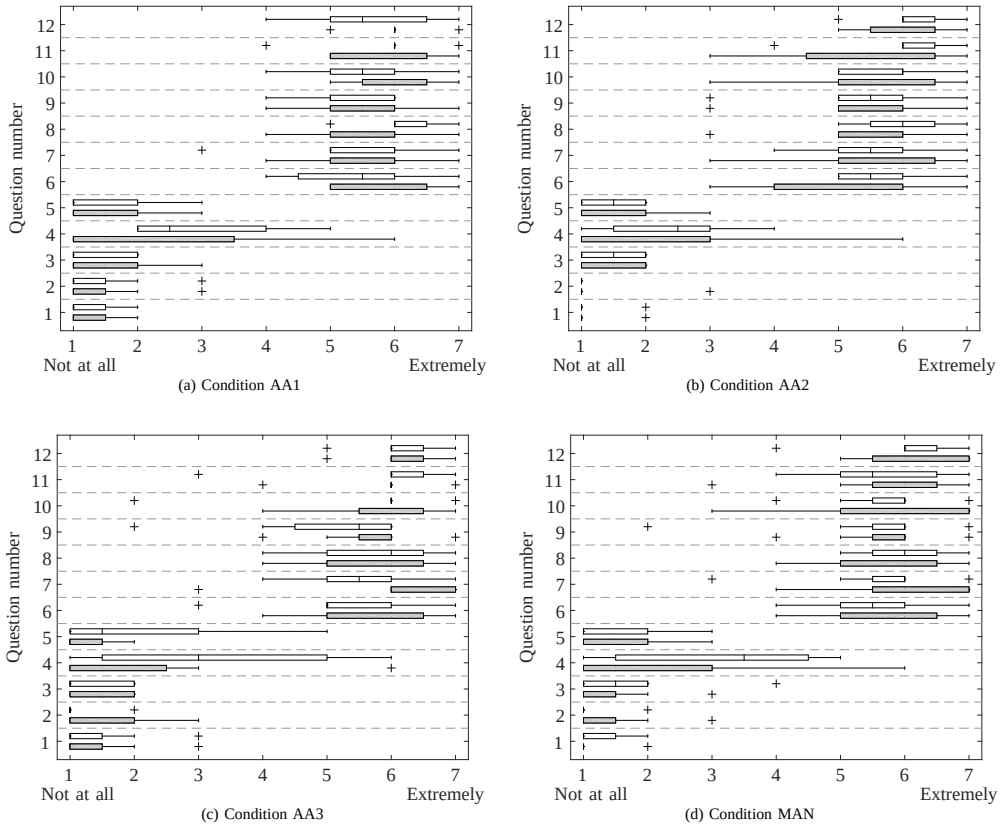


Figure 27: Checklist for trust results for the different automation conditions

No significant difference in both ISA workload and NASA-TLX ratings was found between the Sepmon, PAP and baseline groups. This is not entirely unexpected. As the displays provide the controller with additional information, the conflict detection task should become less demanding and workload originating from frustration was expected to decrease, but the processing of this additional information should increase workload. However, the results from the NASA-TLX also show that there is no difference in both frustration ratings and weights for the separation monitor, projected aircraft path and baseline groups. The frustration questionnaire given at the end of the experiment indicates that the automation is the main reason participants experienced frustration. There might be several reasons for this. At the moment the automation activates, a sound is heard by the participant. This sound was a short low buzzing sound repeated three times. In other situations, this sound is usually interpreted as something being wrong. Therefore, the sound heard upon activation of the automation might cause the participant to think that the automation considers their solution wrong.

It was observed by the experimenter that in the Sepmon and PAP groups, the participants used these functions to try and achieve high efficiency, by trying to achieve a separation between aircraft as close to 5 NM as possible. This sometimes led to the participants activating the automation for a very short term conflict (less than 1 minute ahead). Due to the way the automation constructs its advisories (it does not take the turn rate of the aircraft into account), solutions that did not solve the conflict were suggested by the automation in this case. As some participants had adopted the strategy of accepting all advisories, this led to a few LoSs and even one mid-air collision. Interestingly, this did not lead to a change of the level of trust the participants had in the automation.

It was further noticed by the experimenter that both the Sepmon and PAP groups did not seem to use the changes to the displays upon activation of the automation. Most participants in the Sepmon group did not compare the separations of aircraft pairs between the automation and their own solution,

but reacted to an advisory first and then clicked on the symbols in the separation monitor to identify aircraft pairs that were close to the critical area. Most participants in the PAP group did not use the function to see what the difference was between the automation solution and their own. They used the function to look for aircraft pairs that would conflict in the future and resolve the conflict. When the automation activated, they would first respond to the advisory and then use the projected aircraft path function in the same way they would when the automation was not active.

The results also show that the Sepmon group received more STCAs than both the PAP and baseline groups. A few participants in the Sepmon group mentioned in the end of experiment questionnaire that it was very difficult for them to link the symbols on the separation monitor to aircraft pairs on the PVD. This could mean that the training this group received was not sufficient for them to use the information shown on the separation monitor to its full potential. Furthermore, automation active time was higher for the Sepmon group than for the PAP group, as can be seen in Figure 23, indicating that participants in this group created more induced conflicts when the automation was not active. This too might be related to the apparently insufficient training of the Sepmon group, as there was neither significant difference in automation active time between the PAP and baseline groups, nor between the Sepmon and baseline groups.

Overall, the different kinds of transparency do not appear to have made any difference for the large majority of the recorded variables. When comparing the different kinds of transparency, the projected aircraft path function appears to be more intuitive to use than the separation monitor, as most participants in the Sepmon group reported that it was difficult for them to link the symbols on the separation monitor to aircraft pairs on the radar screen. Both the adapted SASHA questionnaire and the checklist for trust did not show a significant difference between the separation monitor and the projected aircraft path function, however, participants rated both displays favorably.

VII. Recommendations

For future research, the authors recommend that this experiment is performed again, but with an increase in training for the separation monitor group. Experiments for the three different conditions were performed by two different experimenters. Because of this, participants for the different conditions might have received inconsistent information on the simulator and automation. It might be worthwhile to run the experiment with somewhat more experienced controllers, as it is expected that they are quicker in understanding the separation monitor and using its information fully. Furthermore, measurement scenarios with different traffic intensities could be used, in order to see if transparency has more influence on the controller at higher traffic densities.

The Solution Space Diagram (SSD) [3] could be added as another level of transparency, as it contains the same information as the separation monitor and the projected aircraft path function. However, this can only be done in an experiment that allows heading changes only, as the SSD also contains information on the speed envelope of the aircraft, which is not present in the separation monitor and projected aircraft path function. As the SSD can be shown directly on the radar screen, it has more visual momentum than the separation monitor, as the separation monitor is shown on a different screen.

As most participants did not use the information that would show the difference between the automation's solution and their own solution, there might not truly have been different levels of transparency for the different groups. It is recommended that a solution is found where the participants would use this information in their decision making process. However, it is not yet clear in what way this could best be achieved. Furthermore, participants did not seem to be able to separate the automation and the displays. A way should be found to have the controller understand that the displays and the automation are separate entities.

The checklist for trust showed that participants blindly trusted in the automation. Previous research in trust has shown that when controller self-confidence is low, they tend to rely on the automation more than when their self-confidence is high [36, 37]. It might therefore be worthwhile to find a measure for the controller self-confidence. This could for example be adapted from the method used by Lee and Moray, who used subjective questions to measure both operator trust and self-confidence [36].

Improvements to the adaptive automation tool could also be made. The 'negative' sound presented to the controller upon activation of the automation could be replaced by a nicer sound. This could possibly have a positive effect on controller frustration, as the controller might not feel the automation thinks they are wrong when it activates. The box on the top right of the screen holding the advisory expiry time and the accept and reject buttons seemed very disconnected to the participants and it could be considered to display this information in a different way, possibly on the aircraft labels. This should help visual momentum and give controllers slightly more time to consider the advisory. Lastly, the construction of resolution advisories should be changed to incorporate the turn rate of the aircraft, as the current resolution advisories sometimes do not solve a conflict because of this.

VIII. Conclusion

The research presented in this paper looked at the influence of different levels of transparency on controller acceptance of advisories given by and frustration with the adaptive automation tool developed by IJtsma [12]. Two different displays were adapted and used in a human in the loop experiment. The information on the separation monitor is shown in exactly the same way as the automation uses it, while the projected aircraft path function shows the same information in terms of conflicting aircraft pairs and time to closest point of approach, but not in the way the automation uses the information. The separation monitor therefore shows the reasoning of the automation, while the projected aircraft path function does not.

Results show that there is no difference in acceptance and frustration for the different levels of transparency. The

only difference found between the levels of transparency is a slightly higher number of STCAs for the separation monitor group. Workload remained constant across different transparency levels. There was no difference in controller situational awareness between the separation monitor and projected aircraft path groups, nor was there any difference in controller trust in the automation. However, participants from the projected aircraft path group reported that the projected aircraft path function was really helpful, while participants from the separation monitor group reported that they had difficulty linking the symbols on the separation monitor to aircraft pairs on the radar screen.

In conclusion, there does not seem to be any influence of transparency on operator acceptance of the advisories, nor any influence on operator workload, frustration, situational awareness and trust in the automation. However, it is not clear if this influence is absent as a whole or just for this experiment, as participants in the experiment did not fully use the information available to them to try and understand the intentions of the automation. It is therefore recommended that further research on automation transparency in this context is performed, to be able to confirm or dismiss the findings in this paper.

References

[1] Eurocontrol, "Challenges of growth 2013. task 4: European air traffic in 2035," 2013.

[2] A. Majumdar and J. Polak, "Estimating capacity of europes airspace using a simulation model of air traffic controller workload," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1744, no. 1, pp. 30–43, 2001.

[3] R. Göritzlehner, C. Borst, J. Ellerbroek, C. Westin, M. van Paassen, and M. Mulder, "Effects of transparency on the acceptance of automated resolution advisories," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2965–2970, 2014.

[4] SESAR. Sector team operations - en-route air traffic organiser. http://www.sesarju.eu/sesar-solutions/conflict-management-and-automation/sector-team-operations-en-route-air-traffic.

[5] T. Bolic, "Automation adoption and adaptation in air traffic control," Ph.D. dissertation, UNIVERSITY OF CALIFORNIA, BERKELEY, 2006.

[6] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

[7] M. Scerbo, "Theoretical perspectives on adaptive automation," in *Automation and human performance*, R. Parasuraman and M. Mouloua, Eds. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 1996, pp. 37–63.

[8] R. Parasuraman, T. Bahri, J. Deaton, J. Morrison, and M. Barnes, "Theory and design of adaptive automation in aviation systems," Cognitive Science Laboratory, THE CATHOLIC UNIVERSITY OF AMERICA, Washington, DC, Technical Report N62269-90-0022-5931, 1990.

[9] B. Hilburn, R. Molloy, D. Wong, and R. Parasuraman, "Operator versus computer control of adaptive automation." *Seventh International Symposium on Aviation Psychology*, pp. 31–36, 1993.

[10] T. Inagaki, "Adaptive automation: Sharing and trading of control," in *Handbook of Cognitive Task Design*, E. Hollnagel, Ed. Lawrence Erlbaum Associates, 2003, pp. 147–169.

[11] D. Kaber and M. Endsley, "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*, vol. 5, no. 2, pp. 113–153, 2004.

[12] M. IJtsma, "Adaptive automation based on air traffic controllers decision-making," Unpublished M.Sc. Thesis, Faculty of Aerospace Engineering, Delft University of Technology, 2016.

[13] J. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes, "Situation awarenessbased agent transparency," U.S. Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5425, Technical Report ARL-TR-6905, 2014.

[14] M. Ghazizadeh, J. Lee, and L. Boyle, "Extending the technology acceptance model to assess automation," *Cogn Tech Work*, vol. 14, pp. 39–49, 2012.

[15] K. Bessiere, I. Ceaparu, J. Lazar, J. Robinson, and B. Schneiderman, "Social and psychological influences on computer user frustration," in *Media Access: Social and Psychological Dimensions of New Technology Use*, E. Bucy and J. Newhagen, Eds. Lawrence Erlbaum Associates, 2004, pp. 91–103.

[16] A. Maslow, "Deprivation, threat and frustration," *Psychological Review*, vol. 48, pp. 364–366, 1941.

[17] C. Lin, T. Yenn, and C. Yang, "Evaluation of operators performance for automation design in the fully digital control room of nuclear power plants," *Human Factors and Ergonomics in Manufacturing and Service Industries*, vol. 20, no. 1, pp. 10–23, 2010.

[18] D. Kaber, J. Riley, K. Tan, and M. Endsley, "On the design of adaptive automation for complex systems," *INTERNATIONAL JOURNAL OF COGNITIVE ERGONOMICS*, vol. 5, no. 1, pp. 37–57, 2001.

[19] M. Mouloua, J. Smither, D. Vincenzi, and L. Smith, "Automation and aging: Issues and considerations," *Advances in Human Performance and Cognitive Engineering Research*, no. 2, pp. 213–237, 2015.

[20] K. Höök, "Steps to take before intelligent user interfaces become real," *Interacting with computers*, vol. 12, no. 4, pp. 409–426, 2000.

[21] D. Thurman, D. Brann, and C. Mitchell, "An architecture to support incremental automation of complex systems," in *Proceedings of the 1997 IEEE International Conference on Systems, man and Cybernetics*, 1997, pp. 1174–1179.

[22] H. Cramer, V. Evers, S. Ramlal, M. van Someren, L. Ruthledge, N. Stash, L. Aroyo, and B. Wielinga, "The effects of transparency on trust in and acceptance of a content-based art recommender," *User modeling and user-adapted interaction*, vol. 18, pp. 455–496, 2008.

[23] I. Al-Jabri and N. Roztocki, "Adoption of erp systems:

Does information transparency matter?" *Telematics and Informatics*, vol. 32, pp. 300–310, 2015.

[24] A. Roberts and S. Pember, "Air traffic control," Patent US 8,255,147 B2, 2007.

[25] S. B. J. Van Dam, A. L. M. Abeloos, M. Mulder, and M. M. Van Paassen, "Functional presentation of travel opportunities in flexible use airspace: An EID of an airborne conflict support tool," in *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 1. The Hague: IEEE, 2004, pp. 802–808.

[26] International Civil Aviation Organization, "Rules of the Air and Air Traffic Services," Montreal, Tech. Rep. 4444-RAC/501, 1996.

[27] D. Woods, "Visual momentum: a concept to improve the cognitive coupling of person and computer," *International Journal of Man-Machine Studies*, vol. 21, no. 3, pp. 229–244, 1984.

[28] B. Kirwan and M. Flynn, "Investigating air traffic controller conflict resolution strategies," Eurocontrol, Brussels, Technical Report, 2002.

[29] K. Lee, K. Kerns, R. Bone, and M. Nickelson, "Development and validation of the controller acceptance rating scale (cars): Results of empirical research," in *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*, 2001.

[30] S. Brennan, "An experimental report on rating scale descriptor sets for the instantaneous self assessment (isa) recorder," *Portsmouth: DRA Maritime Command and Control Division. DRA Technical Memorandum (CAD5)*, vol. 92017, 1992.

[31] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988.

[32] S. A. Rahman, C. Borst, M. Mulder, and M. Van Paassen, *Measuring Sector Complexity: Solution Space-Based Method*. INTECH Open Access Publisher, 2012.

[33] Eurocontrol, "The development of situation awareness measures in atm systems," Technical Report HRS/HSP-005-REP-01, 2003.

[34] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International Journal of Cognitive Ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.

[35] B. Reeves and C. Nass, "The media equation: How people treat computers, television, and new media like real people and places," *Computers and Mathematics with Applications*, vol. 33, no. 5, p. 128, 1996.

[36] J. Lee and N. Moray, "Trust, self-confidence, and operator's adaptation to automation," *Int. J. Human-Computer Studies*, vol. 40, pp. 153–184, 1994.

[37] S. Lewandowski, M. Mundy, and G. Tan, "The dynamics of trust: Comparing humans to automation," *Journal of Experimental Psychology: Applied*, vol. 6, no. 2, pp. 104–123, 2000.

Part II

Thesis Book of Appendices

Appendix A

Literature study

This chapter will present an overview of relevant literature on several topics. First, several subjects concerning automation transparency will be treated in section A-1. Section A-2 will discuss the effect of transparency on trust and frustration. Finally, section A-3 will cover the use of preview information for increasing transparency.

A-1 Automation transparency

This section will cover several subjects concerning automation transparency. Automation is defined as the execution by a machine of a task that was previously performed by a human operator (Parasuraman, 1997), whilst automation transparency is defined as the level of detail with which the automation communicates the reasoning behind its advice on or solution to a problem to the human operator(Göritzlehner et al., 2014). Subsection A-1-1 will discuss the concept of human-centered automation. In subsection A-1-2, automation transparency and opacity will be discussed. Subsection A-1-3 will cover several automation concepts with relations to human-centered automation and transparency. Subsection A-1-4 will explore some existing Air Traffic Controller (ATCo) automation tools and the transparency of these tools. Finally, in subsection A-1-5, the transparency of the Adaptive Automation (AA) tool of IJtsma (2016) will be discussed.

A-1-1 Human-centered automation

The concept of Human-Centered Automation (HCA) encompasses automation that has been designed to work cooperatively with the human operator when pursuing objectives (Billings, 1997). When multiple persons gather to solve a problem together, they contribute different perspectives and different areas of expertise. Different perspectives encompasses different goals, different background knowledge and different assumptions. Different areas of expertise usually mean that the individuals will use different languages in talking about the problem. This makes solving a problem together both beneficial and difficult. On one side, the

resources used in solving the problem are increased, which means that the different perspectives and skills will help overcome some of the possible shortcomings of an individual. On the other side, there is a coordination cost. The diversity that is advantageous when solving the problem can be disadvantageous in coordinating the collaborative problem solving (McCarthy et al., 1991). This same concept can be extrapolated to the joint problem solving of a person with automation. Automation that gives the operator information on automation modes, system states, and future automated actions, particularly if it does so with good etiquette, can improve human-automation communication and therefore possibly increase performance (Sheridan & Parasuraman, 2005).

In order to complement the human operator and minimize the coordination cost, the machine must allow the human operator to easily perform the following actions (Inagaki, 2008):

- Understand what the machine can and can not do.
- Give directives to the machine.
- Monitor what the machine is doing.
- Intervene in machine control when necessary.

The human operator needs to be able to maintain an understanding of the problem from the machine's perspective. In order to be able to do this, the machine needs to provide adequate feedback. Observability of the machine's activities is not enough for making them team players; the human operator must also be able to redirect the machine at any time. In this way, the human operator can acquire control by giving up some, but not necessarily all, potentially useful automation functions. Another reason important reason for directability is the fact that unanticipated situations will arise in complex work environments. This requires human-machine systems to be able to adapt (Mercado, 2013; Christoffersen & Woods, 2002; Woods & Sarter, 1998).

Many human-machine interaction problems are caused by the internal logic of the machine, not the design of the user interface. When automation's functionality and logic are not limited to a clearly defined, restricted set of rules, this makes it difficult for the operator to develop a good comprehension and proper expectations of the automation's behavior. Combining this with a compact and uniform mental model that closely matches the operator's task requirements offers the best possibility that the automation will be simple in to learn and use (Riley, 1996).

A-1-2 Automation opacity/transparency

Automation transparency is defined as the level of detail with which the automation communicates the reasoning behind its advice on or solution to a problem to the human operator(Göritzlehner et al., 2014). Opacity is the opposite of transparency. High opacity can cause the operator to be unable to develop a good mental model of the automation. However, some level of opacity is always needed in order to avoid overloading the operator with information and data (Mouloua et al., 2015).

Höök (2000) has identified three distinct types of transparency, namely domain transparency, internal transparency and embedding transparency. Automation that features domain transparency allows the operator to see through the automation and look at the domain behind it. Automation that features internal transparency allow the operator to see through the automation's interface and observe the internal workings of the automation. Embedding transparency refers to transparency of the entire environment in which the automation will be used.

A way to reduce automation opacity is to make the automation 'think' like the human operator about activities and let the automation provide continuous and detailed feedback on its activities. This means that the structure of the automation and its process knowledge should be compatible with that of the human operator(Thurman et al., 1997).

Previous research has shown that an explanation of the advice of an art recommendation system increased the acceptance of the recommendations. However, an increase in trust in the system was not found (Cramer et al., 2008). Furthermore, Al-Jabri & Roztocki (2015) have found a direct significant effect of transparency on perceived usefulness and ease of use of enterprise resource planning systems. This research also showed an indirect relationship of transparency on the attitude toward using these systems.

A-1-3 Automation concepts

This section covers several automation concepts which can be used in combination with human-centered automation and can feature automation transparency.

Levels of Automation

Different systems rely on automation or human operator in varying degrees. In order to classify the degree of automation of a system, several researchers have created a Level of Automation (LOA) taxonomy. The LOA taxonomy of Sheridan & Verplank (1978) has 10 distinct levels and can be seen in table A-1.

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 informs 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 and it, 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.

Table A-1: LOA taxonomy by Sheridan & Verplank (1978)

This LOA taxonomy has, however, been formulated to apply in the field of teleoperations. Although an attempt was made to generalize, it might not apply in all different situations in which automation is used. Endsley & Kaber (1999) have created a LOA taxonomy with 10 levels that is intended to apply to a broader scope of cognitive and psychomotor tasks that require real-time control and occur within numerous domains. This LOA taxonomy can be found in table A-2 and more detailed descriptions of the different LOAs can be found in Endsley & Kaber (1999).

1	Manual Control (MC)
2	Action Support (AC)
3	Batch Processing (BP)
4	Shared Control (SHC)
5	Decision Support (DS)
6	Blended Decision Making (BDM)
7	Rigid System (RS)
8	Automated Decision Making (ADM)
9	Supervisory Control (SC)
10	Full Automation (FA)

Table A-2: LOA taxonomy by Endsley & Kaber (1999)

Parasuraman et al. (2000) argue that not all parts of a single process should necessarily have the same LOA. They divide the process in four stages, namely:

1. Information acquisition
2. Information analysis
3. Decision selection
4. Action implementation

Each of these different stages in a process can have a different LOA, independent of the others.

Research has shown that LOAs involving the computer performing the implementation phase of a task improved the operator/system functioning. However, systems using a LOA that require both the human and the computer to generate solution options show a decrease in performance when compared to purely human generation of solutions. It appears that humans benefit most from physical implementation assistance and might be hindered in their performance when assistance is provided with higher level cognitive functions (Endsley & Kaber, 1999). A possible explanation for the decrease in performance involving LOAs with computer generated solution options is that the operator needs to spend time trying to understand the computer solutions, especially when the automation is not very transparent.

An extensive review of previous experiments concerning different LOAs by Onnasch et al. (2013) shows that routine performance of a task increases with increasing LOA. However, when looking at failure of automation, it is seen that failure performance decreases with increasing LOA. A slight trend showing decreasing Situational Awareness (SA) with increasing LOA was also observed, but this was found not to be significant. These results indicate that

when automation can not be made 100% reliable, a careful trade-off has to be made between increased routine performance and decreasing failure performance.

Typically, automation develops its own solution to the problem it was designed to (help) solve. Operators find it difficult to accept such advice without some information about how the automation arrived at these conclusions. This led many developers of automation to generate ways to represent knowledge in such automation so they could provide the operator with a description of how the automation generated its solution. However they were generated and however they were shown to the user, these explanations were only shown at the end of the problem-solving of the automation, after the automation had arrived at its solution. As a result, these explanations were one-shot, retrospective explanations for activity that had already been performed. The explanation is then presented at a time when the operator is likely to be busy with multiple activities as a consequence of the cascade of effects of the initial event and increasing cognitive demands to understand the situation and react as this situation evolves (Woods & Patterson, 2000).

The problems described before might be the result of inappropriately designed automation. The operations are performed appropriately under normal conditions. However, when there is inadequate feedback and interaction with the humans who must control the overall conduct of the task, this leads to difficulties for the human controller when the situations exceed the capabilities of the automatic equipment (Norman, 1990).

Adaptive Automation

Adaptive Automation (AA) is a form of automation in which the amount and kind of automation support given to the user can vary between different levels. The decision to regarding the use, disuse and level of support can be made by either the human, the machine or both, depending on the system. Adaptive automation can adjust operator support based on dynamic and situational demand (Scerbo, 1996; Bubb-Lewis & Scerbo, 1997). Different tasks may be completed by the human operator or the computer, depending on the situation.

The main argument against leaving the human operator in charge of changing the amount of support given is that they are probably too busy to make the change when they need it. Operators may be unprepared to manage automation when abrupt, unexpected changes in workload occur, which leaves them to manage a system that they are unable to handle without adequate machine support (Scerbo, 1996).

When the machine has the authority to switch between the different levels of adaptive automation, different triggers for such a switch can be used. Typical triggers can be divided into four categories (Parasuraman et al., 1990; Hilburn et al., 1993; Inagaki, 2003; Kaber & Endsley, 2004):

- **Critical-event logic**: Automation adapts based on the occurrence of a specific event.
- **Dynamic assessment of operator mental workload**: Automation adapts based on an assessment of operator mental workload determined in real-time.
- **Dynamic psychophysiological assessment**: Automation adapts based on psychophysiological measurements (e.g. heart rate, pupil dilation, EEG).

- **Operator performance models**: Automation adapts based on features of a given pilot performance model.

If good performance models are available, it may be possible to extract 'leading indicators'. Leading indicators are precursors that, when observed, indicate the occurrence of some subsequent event (Inagaki, 2003). The AA trigger of IJtsma (2016) is theorized to be such a leading indicator.

People construct and maintain a mental model of the system with which they interact. The model is constructed and updated based entirely on feedback from the system, information from the environment and their own instructions. However, in modern systems, this model depends critically upon the information displays (Norman, 1990).

Care must be taken to keep the human operator involved in the system and allow them to construct and maintain a mental model. Several studies have shown that, when only monitoring an automated system, operators often are slow in detecting problems in which operator intervention is necessary. It has furthermore been shown that additional time is also needed for the operator to reorient themselves to the current state of the system and obtain sufficient understanding of what is happening in order to act appropriately. This is called human Out-of-the-Loop (OOTL) performance (Endsley, 1996; Kaber & Endsley, 2004).

According to Norman (1990), continual feedback on the state of the system is needed, in a natural way, much in the same way as human participants will discuss the issues among themselves when participating in joint problem-solving. This means designing systems that are informative, yet non-intrusive, so the interaction is normal, continuous and the amount and form of feedback adapts to the interactive style of the participants and the nature of the problem. However, this increase in transparency can lead to an increase of operator cognitive workload, as all this information has to be interpreted (Miller, 2014).

Intelligent Tutoring Systems

An Intelligent Tutoring System (ITS) is a computer based instruction system in which some form of Artificial Intelligence (AI) is implemented. The two main places for intelligence in an ITS are in the knowledge the system has of its subject domain and the principles by which the system tutors and in the methods by which these principles are applied. (Anderson, 1988).

Working memory stores what the operator currently knows about the problem. The capacity of this working memory is assumed to be limited. It is possible that the capacity required for the solution of a particular problem will be exceeded and thus that critical information is lost. This can result in a failure to solve the problem or errors in solving the problem. Many errors of learners are due to failures of working memory rather than to failures of understanding (Anderson et al., 1985).

Based on an internal model, the ITS determines the correct solution to a problem and compares this to the student's solution. If the student response is correct, the tutor is quiet, but when the solution is incorrect, the tutor infers the most likely mistake the student made and gives feedback to the student with appropriate remedial instruction (Anderson et al., 1985). One of the issues in the design process of an ITS is when it should provide feedback and in what form. Corbett et al. (1997) propose the following guidelines for feedback timing:

1. Feedback should be provided at a time when the relevant information can be communicated effectively to the student
2. Feedback should not change the requirements of the task
3. Feedback should not disrupt performance of the task
4. To optimize learning of the target skill with respect to time, feedback should be presented as early as possible

The content of feedback is another important factor. What a tutor says to a student is critical and it is helpful to say more than just 'right', 'wrong' and 'do this' (Anderson, 1988). It is therefore necessary to know exactly what mistake the student made and what feedback will help them understand their mistake. In other situations, when the student does not know what to do next or their behavior matches no solution step in the program, the student might be confused. The best thing to do in such situations is to tell the student what to do next. If explained properly, the student is often able to continue on a right track again (Anderson et al., 1985).

A-1-4 Automation transparency in Air Traffic Control

This subsection will discuss several current Air Traffic Control (ATC) automation systems and their individual levels of transparency.

Solution Space Diagram

The Solution Space Diagram (SSD) is a decision-support tool for ATCos developed by Delft University of Technology. In its simplest form, the SSD gives a visualization of conflicting and non-conflicting areas in both speed and heading for an aircraft and can be seen in figure A-1. This display helps controllers resolve traffic conflicts, because they can avoid conflict areas in the aircraft's solution space. The SSD features automation transparency, as it calculates which headings and speeds would result in a conflict between two aircraft and displays it to the controller. However, it is not completely transparent, as the SSD does not show how the forbidden zones have been calculated.

Göritzlehner et al. (2014) investigated the influence of transparency on the acceptance of resolution advisories generated by automation. They used the SSD to present these automated resolution advisories within the solution space. The results of this experiment showed no significant effect of transparency on acceptance of advisories.

En Route Air Traffic Organiser

En-route Air Traffic Organiser (ERATO) is an ATCo tool developed by Single European Sky ATM Research (SESAR). It includes medium term conflict detection (MTCD) and a first step towards a conflict resolution assistant and aims at making air traffic and time management easier through four main features: filtering, task scheduling, extrapolation and geographic markers (SESAR, n.d.).

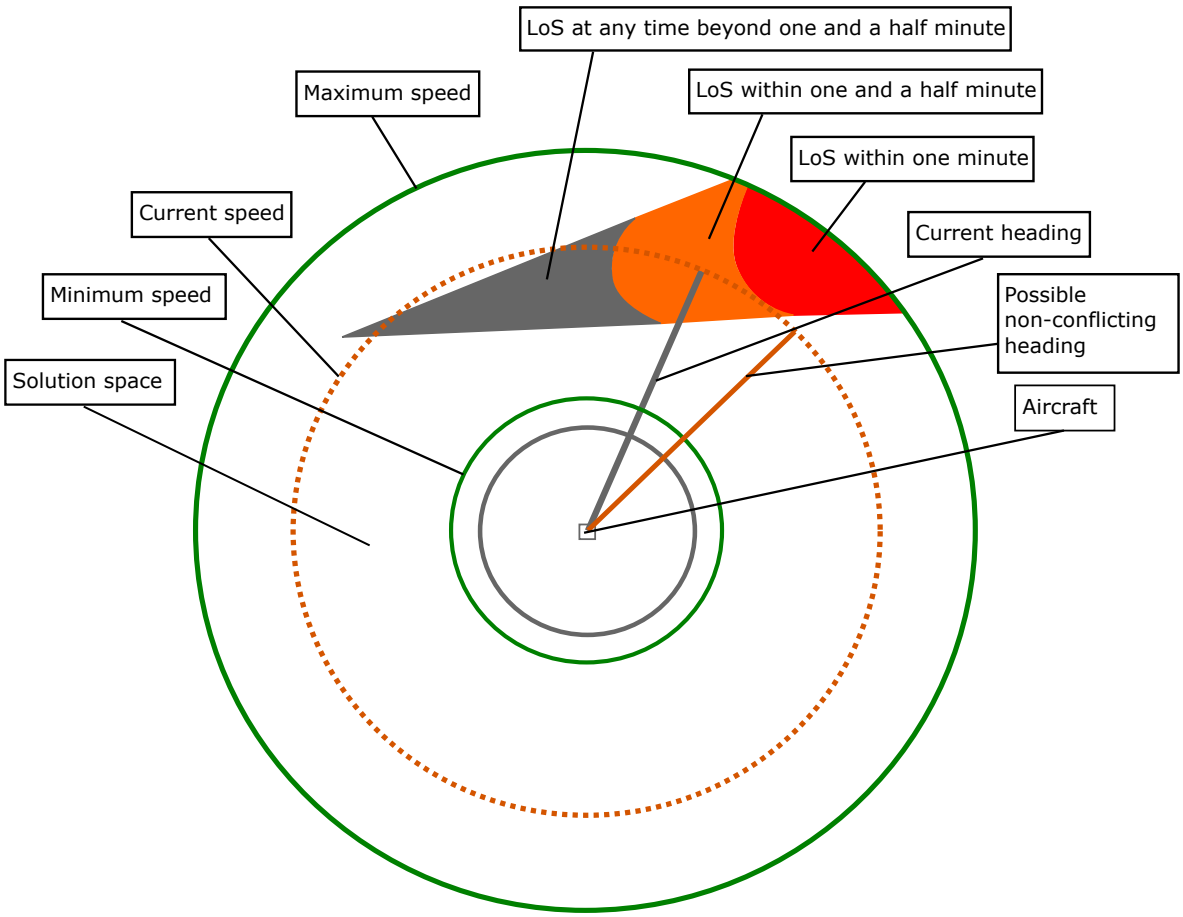


Figure A-1: The SSD as used by Göritzlehner et al. (2014).

If the controller needs to analyze a traffic situation in more detail, they can use the filtering function to isolate aircraft that could experience a Loss of Separation (LoS). From this filtering mode, the controller can use the extrapolation function to extrapolate aircraft position in relation to control parameters. ERATO displays the future positions of each of the conflicting aircraft. ERATO also highlights flights that diverge from their ATC assigned trajectories. In the task scheduling function, conflicts are shown on a time line. The controller can isolate a group of aircraft in order to find the best scheduling solution for them(DSNA & ENAV, n.d.).

The interface the controller is presented with in ERATO can be seen in figure A-2.

The Medium-Term Collision Detection (MTCD) tool alerts the ATCo up to 20 minutes before a predicted conflict. In order to do so, the MTCD tool uses a trajectory prediction algorithm based on planned and tactical trajectories to determine likely future aircraft positions. If a LoS is found in these predicted positions, the ATCo will receive a warning (Eurocontrol, 2010). The tool features automation transparency through the extrapolation function. This function allows the ATCo to have the ERATO tool to show the predicted trajectory for the conflicting aircraft on the radar screen, allowing them to see why the automation gives a MTCD warning.



Figure A-2: The controller screen using ERATO (ENAV, n.d.).

User Request Evaluation Tool

In order to keep track of flights in their sector, ATCos use Flight Progress Strips (FPSs). A FPS is a rectangular piece of paper on which detailed flight information is displayed. ATCos update this strip by writing on it, by placing it in a certain location among the other FPSs, or by offsetting it from other FPSs (Edwards et al., 1995).

ATCos in the United States have a tool called User Request Evaluation Tool (URET) at their disposal. URET has several functionalities that can be used independently of each other. The intended purpose of URET is supporting strategic planning. The core functionality of URET is a MTCD tool, but the tool also includes electronic FPS replacement, a trial planning function and a route amendment function. Figure A-3 shows the Aircraft List and Plan display that is part of URET. Potential conflicts are shown in this Aircraft List and Plan display, as well as trial planning results (Bolic, 2006).

The ATCo can activate a Graphic Plan Display on which they may graphically display aircraft trajectories and conflict situations, allowing them to see the cause of a MTCD alert, but also use this for trial planning. Bolic (2006) researched the use of URET in different control centers and concluded that URET is not always used for its intended purpose. Controllers indicated that they often noticed conflicts ahead of the MTCD functionality. Overall, it seemed that the system is used more as an electronic FPS replacement than anything else, despite the original purpose of the tool.

The system does not feature transparency in the MTCD tool, as it only indicates conflicts, but does not indicate why there is a conflict. The route amendment function does feature transparency, as it allows the controller to directly plan a flight on a map display which shows the domain of possibilities.

A-1-5 The transparency of the AA tool of IJtsma (2016)

For each conflict, the AA tool of IJtsma (2016) considers four possible solutions, namely steering each individual aircraft to the left or right. An example of this can be seen in figure A-4.



Figure A-3: The Aircraft List and Plan display in URET (Bolic, 2006)

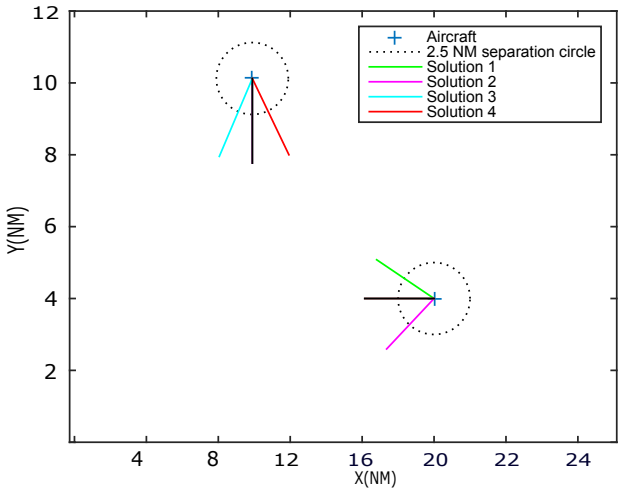


Figure A-4: The solutions considered by the AA algorithm (IJtsma, 2016).

The best solution among these four possible solutions is determined by the minimization of a cost function. This cost function encompasses several different variables, among which are the number of induced conflicts, the heading deviation and the airspeed. Compared to the other variables in the cost function, the number of induced conflicts has a weighing factor more than an order of magnitude greater than that of any other variable. This ensures that

the number of induced conflicts will always be minimal, which was the purpose of creating the automation.

Three different triggering thresholds were tested by IJtsma (2016), namely one, two or three bad decisions made by the controller. A bad decision was counted as a controller resolution of a conflict in which the controller resolution had a higher number of induced conflicts than the optimal solution as determined by the optimization algorithm. Upon triggering the high LOA, the aircraft with which a conflict would have been created had the controller solution been implemented are briefly highlighted in blue. After triggering of the automation, the automation gives resolution advisories for all conflicts occurring within five minutes. When a resolution advisory is proposed to the operator, they have a window of fifteen seconds in which to reject or accept the advisory. If no action is undertaken in this window, the resolution advisory is implemented automatically. The advisories are shown in an area to the right of the Plan View Display (PVD). The PVD is the radar screen that shows the location and movement of the aircraft in the sector. A schematic of the user interface of the AA tool can be seen in figure A-5.

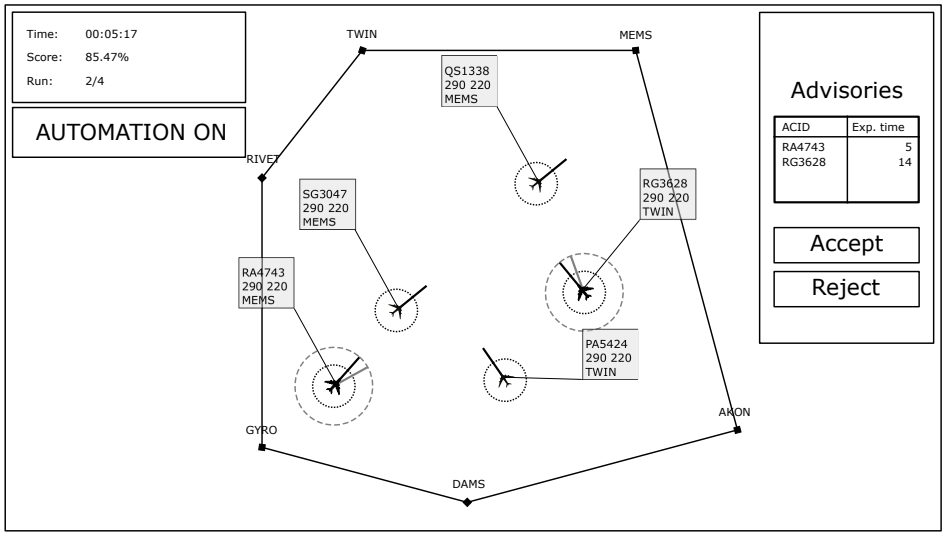


Figure A-5: Schematic of the user interface of the AA tool of IJtsma (2016). Dimensions are not to scale.

When an advisory is clicked on, the aircraft the advisory pertains to get highlighted. When the interface is looked at, it is thus transparent in the sense of showing the time left to react to an advisory and which conflict the advisory is meant to solve. However, there is little to no transparency in terms of showing the inner workings of the AA. The controller is not informed on the workings of the automation, both in the pre-experiment briefing and during the experiment, and thus has no way to construct a mental model of the workings of the automation. The only information the controller receives about the automation is that it focuses on preventing the controller from inducing additional conflicts when implementing a resolution. This is shown when the aircraft with which a conflict would have been created had the controller solution been implemented are briefly highlighted in blue upon activation of the high LOA. However, during the time resolution advisories are given, no such information is available.

A-2 Effect of transparency on acceptance and frustration

This section will discuss the effect of transparency on acceptance and frustration. In subsection A-2-1, human trust in automation will be discussed. Subsection A-2-1 will treat the subject of human frustration with automation.

A-2-1 Building trust through transparency

According to Lee & See (2004) trust can be defined as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability". As with humans, people develop trust relationships with automation, relying on automation they trust and rejecting automation they distrust. As the human operator will ultimately be responsible for the system performance, trust in the automation is an important factor in the use of automation tools (Lee & See, 2004).

A lot of problems arising from the interaction of human and automated decision aid stem from inappropriate expectations of the capabilities of the system (Mosier & Skitka, 1996). Proper trust calibration is important, because overtrust in automation may lead to the operator to follow unreliable automation, even when this conflicts with other information available to the operator (Wickens, 1998). Undertrust in automation may lead to disuse of a helpful tool, which leads to the loss of its potential benefits to system performance. To avoid problems of over- and undertrust, the operator's trust must be calibrated to the automation (Muir, 1987).

Early in their experience, an operator will judge the predictability of automation by evaluating the consistency of the automation's recurrent behaviors. The growth of trust depends on the human operator's ability to make a good estimate of the predictability of the machine's behaviors. Unless the automation's behavior is completely deterministic, these behaviors will be distributed with a variance about some mean. The operator's ability to estimate these depends on their own limitations as a decision maker and on certain properties of the automation and its environment. As an operator's trust in automation is based on their observations of the automation's behavior, this behavior must be observable for the operator's trust in the automation to grow (Muir, 1987).

Strategic conformance is the degree to which the automation's behavior and apparent underlying rationale match those of the operator. Westin et al. (2015) have shown that when a decision support tool provides conformal advisories, they are more often accepted by the operator and are also rated higher and responded to faster. Conformance enables the operator to generate a good estimate of the machine's behavior, as this is compatible with their own problem-solving method and should therefore promote proper trust calibration.

Another important factor related to human trust in automation is self-confidence. When an operator's trust in the system exceeds their self-confidence in manually performing a task, they tend to use the automation. When the operator's self-confidence exceeds their trust in the automation, they will opt for manual control where possible. Human subjects often tend towards overconfidence. This, combined with possible miscalibrations of trust points towards potential problems in the operator's cooperation with automation (Lee & Moray, 1994; Lewandowski et al., 2000).

Past studies have shown that humans sometimes question the accuracy and effectiveness of the automation's actions or advice due to the their difficulties understanding the state of the

automation and the reasoning behind the behaviors (Chen et al., 2014). If the automation's actions are comprehensible to the operator and the reasoning behind the decision is transparent, the operators may view the automation more favorably (Ghazizadeh et al., 2012). One way to achieve this and proper trust calibration is to make the automation's algorithm transparent to the user. When this is done, the operator can develop a more robust mental model of the automation's workings and this leads to more appropriate usage of the automation (Chen et al., 2014; Oduor & Wiebe, 2008; Norman, 1990).

A-2-2 Decreasing frustration through transparency

Frustration arises when there is a hindering condition that hampers or prevents the attainment of a goal (Bessiere et al., 2004). However, according to Maslow (1941), a distinction can be made between deprivation and threat. There is a difference between a deprivation that is not important to the individual, something that is easily replaced or which has few serious consequences, and a deprivation that is a danger to the person, their goals in life, their defensive system, their self-confidence or their sense of security. The effect of frustration can also be found as a result of other types of threat, conflict, rejection, etc. (Maslow, 1941).

Responses to frustration by humans can be either adaptive or maladaptive. Adaptive responses are constructive and are aimed at solving the problem that is preventing the goal attainment. Maladaptive responses consist of the absence of constructive problem-solving and usually create additional problems, thereby making the frustrating experience worse. Factors that influence the level of frustration experienced by users of computer systems include the level of dedication to the goal, the severity of the disruption and the strength of will to obtain the goal. A major cause of user frustration with computers can be attributed to bad or confusing interface design (Bessiere et al., 2004).

Human frustration with automation is not new; as long as there has been automation, there have been humans frustrated with some factor of this automation. Lin et al. (2010) performed an experiment in which the participants had to perform a nuclear plant shutdown task in either a manual or an automated mode. The results of this experiment show that significantly higher levels of frustration occurred in participants under the automatic control mode. In the open questions, a participant stated that they felt OOTL during the shutdown task in automated mode and that this may have been the cause of the frustration of this particular participant. Kaber et al. (2001) state that an increase of task requirements beyond the human operator's cognitive capabilities can lead to a feeling of frustration and defeat, as well as leading to a loss of self-confidence in the ability to perform the task.

Increasing system transparency might prevent the operator perceiving the automation as trying to inhibit or prevent the realization of their goal, especially when this prevents a conflict between the operator and the automation. Furthermore, increased transparency may encourage adaptive responses whenever frustration does occur.

A-3 Using preview information to achieve transparency

In the domain of human manual control, the ability too 'look ahead ' or preview is often important. Experiments have shown that performance improved with increasing preview

time. However, preview is only useful up to a certain point into the future. Beyond this point usefulness of the preview decreases (Tomizuka & Whitney, 1976). The 'tunnel-in-the-sky' display for pilots is an example of such a preview display. It shows the future path of the aircraft as a series of squares of decreasing size, which is seen as a tunnel through which the aircraft flies. Grunwald et al. (1980) have shown that adding predictor/director symbols to this display increases performance for helicopter pilots for both trajectory following and trajectory entry.

It has also been shown that aircraft pilots benefit from preview information under self-separation conditions. Providing the pilot with the intruder's predicted flight path next to his own predicted flight path decreased workload and improved safety (Wickens & Morphew, 1997). The usefulness of preview in ATC when dealing with traffic under self-separation has also been shown. With a preview display, controllers completed more flights and seemed better at correcting their errors (Endsley et al., 1999).

The SSD, discussed before in section A-1-4, also provides the ATCo with some form of preview information. The conflicting areas in speed and heading are constructed under the assumption that all aircraft in the sector maintain their current heading, speed and altitude.

An effect of preview has also been found in the current ATC system. Klomp et al. (2011) found that including a preview functionality in the plan view display did not result in an increase in SA, but it did result in lower workload for the ATCo. Furthermore, all participants partially or fully agreed that the preview increased their insight in the traffic and that the preview works intuitively.

Appendix B

Display adaptations

This chapter will discuss the approach used to increase the transparency of the AA tool of IJtsma (2016). Section B-1 will show the workings of the AA tool. In section B-2, the separation monitor will be introduced and in section B-3, it is shown why the separation monitor can be used to increase the transparency of the AA tool. Section B-4 will discuss the adaptations that have been made to the separation monitor. Finally, section B-5 will introduce the projected aircraft path functionality.

B-1 Workings of the AA tool of IJtsma (2016)

Figure B-1 shows the flow diagram for the software of the AA tool. At every point in time, the automation maintains and updates a separation monitor. The automation uses the separation monitor data to determine which aircraft pairs will result in a conflict.

Whenever the controller selects an aircraft, if this aircraft has a future conflict with another aircraft, the automation will create a resolution advisory for this aircraft pair. As mentioned in section A-1-5, the automation considers four possible solutions. When considering a solution, the automation adds an extra separation margin to the solution, to allow for some operator inaccuracies in directing aircraft on a computer screen.

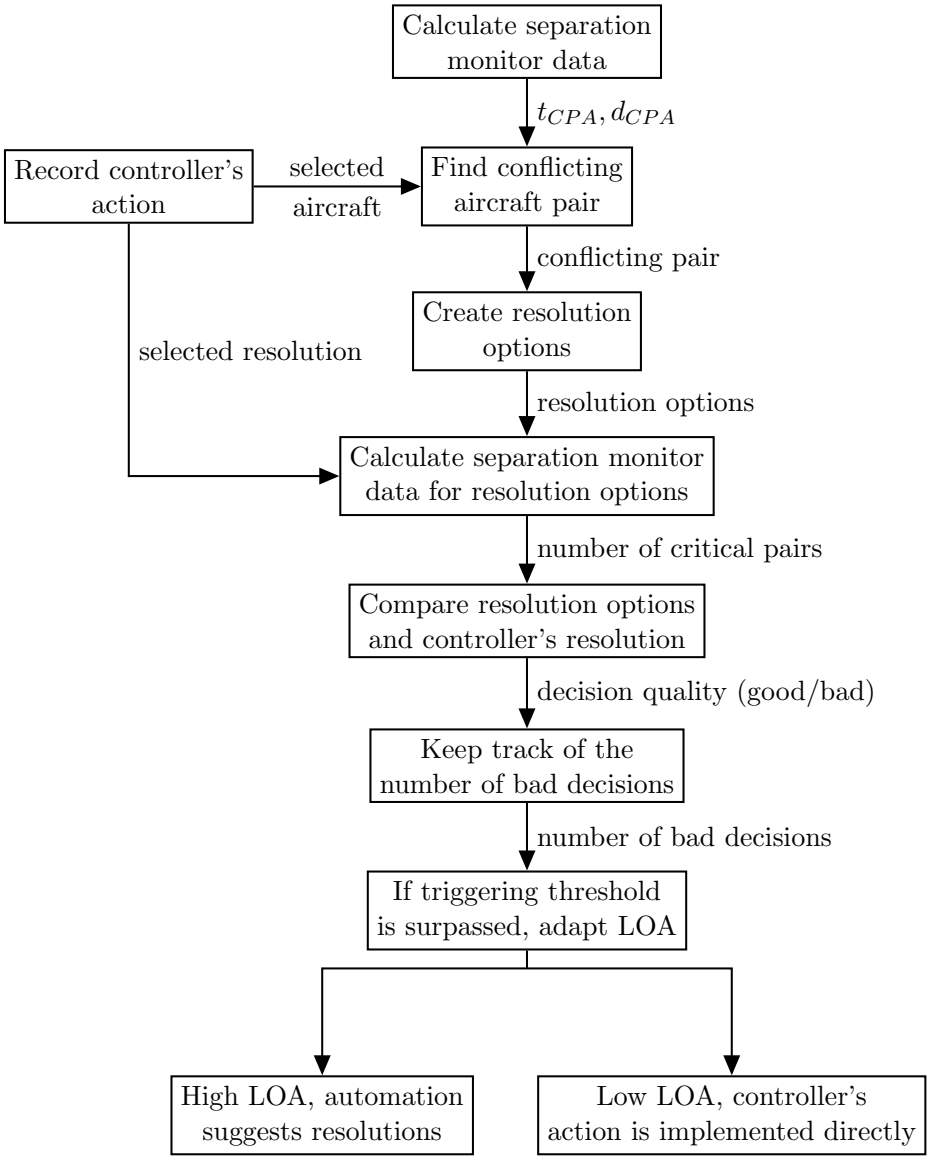
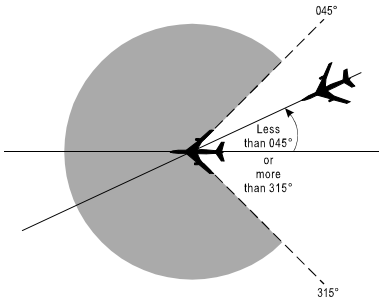


Figure B-1: Flow diagram for the AA tool of IJtsma (2016)

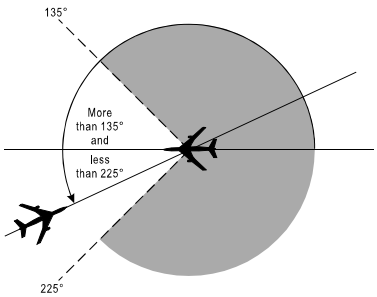
As different types of conflicts can have different optimal solutions, IJtsma (2016) divided conflicts over three different categories.

- Same track conflicts (less than 45 or more than 315 degrees heading difference)
- Reciprocal tracks conflicts (more than 135 and less than 225 degrees heading difference)
- Crossing tracks conflicts (heading difference between 45 and 135 or 225 and 315 degrees)

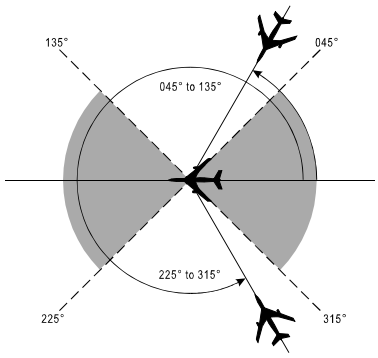
The different conflict geometries can also be seen in figure B-2.



(a) Same track



(b) Reciprocal tracks



(c) Crossing tracks

Figure B-2: Conflict classification as used by ICAO (International Civil Aviation Organization, 1996)

IJtsma (2016) identified that for crossing conflicts ATCo generally steer the slower aircraft behind the faster aircraft. However, for same track conflicts, the faster aircraft is often vectored direct to route, in front of the slower aircraft. A standard solution for reciprocal track conflicts was not found. The automation uses this information when determining the best solution to a conflict.

The automation constructs a cost function for each of the four possible solutions. This cost function consists of five factors:

1. The change in the number of aircraft in the critical area of the separation monitor (ΔN)
2. Time to Closest Point of Approach (CPA) (t_{CPA})
3. Heading deviation ($\Delta\Psi$)
4. Separation at CPA (d_{CPA})
5. Speed (V)

For the different categories of conflicts, different weights are used for the factors in the cost function and these can be found in table B-1.

Table B-1: Weights for each factor in the cost function in the AA tool of IJtsma (2016)

	Factors				
Conflict type	ΔN	t_{CPA}	$\lvert\Delta\Psi\rvert$	d_{CPA}	V
Same track	10	0	0.2	0	-0.02
Crossing	10	0.65	0.02	-0.2	0.01
Reciprocal	10	0.55	0.1	0	0

As the purpose of the automation is to prevent excessive workload by preventing or minimizing the number of induced conflicts, the change in the number of conflicts has a weight that is more than an order of magnitude larger than that of all other factors in the cost function. The best solution to the conflict is found by minimizing this cost function.

An updated separation monitor is then calculated for both the controller solution and the automation solution and the results are compared. If the controller solution has more induced conflicts than the automation solution, this is recorded as a 'bad' solution. The automation keeps track of the number of 'bad' solutions and will switch to the high level of automation if the triggering threshold is surpassed. When the high level of automation is active, the automation actively provides its solutions to the controller, who can either accept or reject the solution within fifteen seconds. When there is no response from the controller in these fifteen seconds, the solution will automatically be implemented.

B-2 Separation monitor

The separation monitor is an ATC system that consists of a trajectory prediction module and a conflict detector. The trajectory prediction module calculates a trajectory for each aircraft and inputs aircraft detected position data. The conflict detector uses this information to

detect future circumstances under which pairs of aircraft violate the separation minimum and shows this information on a display device. The system uses two proximity tests. The second one is more restrictive than the first one. The results of these proximity tests are displayed using a symbol representing each pair of aircraft which violate the second test in a first mode, and which violate the first test, but not the second test, on a second mode. Figure B-3 shows the radar screen the ATCo sees, called the Plan View Display (PVD), and the corresponding separation monitor display (Roberts & Pember, 2007).

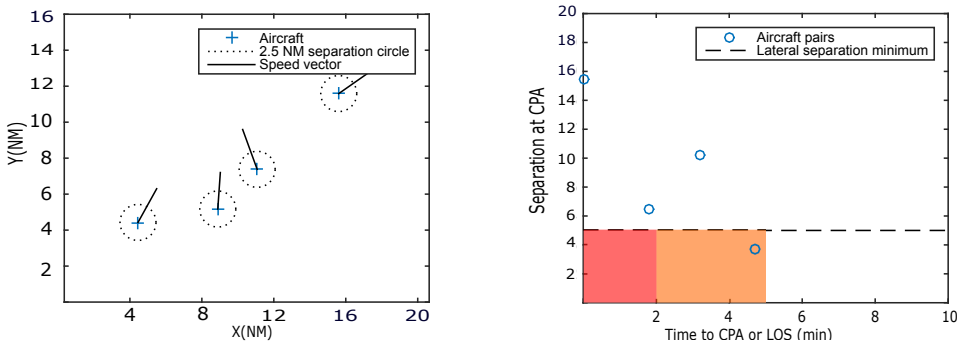


Figure B-3: The PVD and the corresponding separation monitor (IJtsma, 2016).

In figure B-3, each circle represents a pair of aircraft. When an aircraft pair appears below the dashed line, it will violate the separation minimum at some time in the future, which is indicated on the x-axis, with the minimum distance between the aircraft indicated on the y-axis. When the aircraft pair will violate the separation minimum at some point in the future, the distance and time indicated are the distance at CPA and time to LoS. When the aircraft pair will not violate the separation minimum at some point in the future, the distance and time indicated are the distance at CPA and time to CPA. The red area indicates the occurrence of a Short-Term Collision Alert (STCA) for a particular aircraft pair and the orange area indicates the look-ahead time of the adaptive automation developed by IJtsma (2016). The red and orange areas combined constitute the critical area of the separation monitor.

B-3 Increasing transparency

In order to increase the transparency of the AA tool developed by IJtsma (2016), an adaptation of the separation monitor as developed by NATS has been used (Roberts & Pember, 2007). The conflict detection and resolution capabilities of the AA tool have been based on the separation monitor, which makes it the perfect candidate for showing the inner workings of the tool.

In the cost function used by the automation to determine the best solution to a conflict, the change in the number of conflicts, or the number of induced conflicts, has a weight of more than an order of magnitude greater than any other parameter. When some adaptations are made, the separation monitor is capable of providing a natural representation of this number of induced conflicts.

Showing this largest influence on the decision of the AA increases the observability of the automation. Because the tool features management by exception, the AA should already be directable. Furthermore, the presence of information on the number of induced conflicts should show the consistency of the tool's advice in preventing induced conflicts, promoting growth of trust in the automation. As frustration can be caused by conflict, showing the reasoning behind the AA's logic should decrease the feeling of conflict with the automation, because there is a clear explanation for it's logic. Even though an operator can still disagree with the automation's advice, the presence of an explanation for the advice should decrease the feeling of conflict in the operator in most cases.

B-4 Separation monitor preview functionality

In order to help the operator understand the reason for the automation's advice, a preview function has been implemented in the separation monitor. This function shows a preview of the separation monitor when the resolution advisory would be accepted and another preview of the solution the controller suggests. In this way, the controller can see the result of their actions and the results of the resolution advisory and compare them. The preview functionality of the separation monitor can be seen in figure B-4.

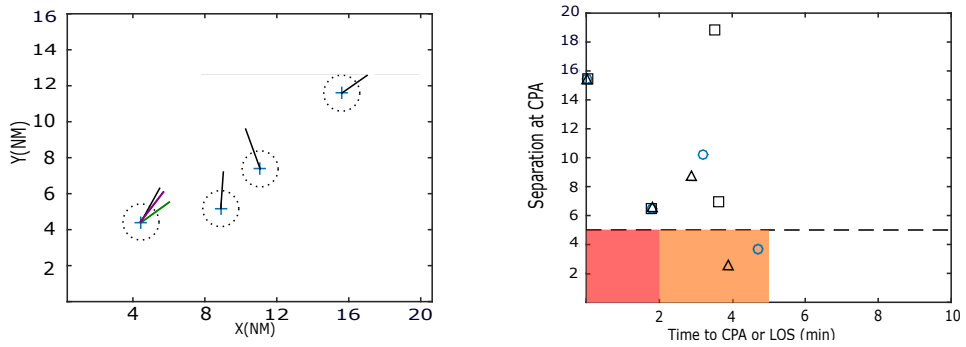


Figure B-4: The PVD and the corresponding separation monitor with preview functionality.

Figure B-3 shows the separation monitor without previews and figure B-4 shows the separation monitor with both previews. In these figures, the circles represent conflicting aircraft pairs in the current situation. The squares represent the conflicting aircraft pairs when the solution proposed by the automation would be implemented. The triangles represent the conflicting aircraft pairs when the solution proposed by the operator would be implemented. In the PVD, the purple line represents the solution suggested by the controller and the green line represents the automation resolution advisory.

The choice has been made to use triangles and squares to indicate the different sets of data in the separation monitor, as opposed to using circles in different colors. As possible participants might suffer from varying degrees of colorblindness, these people would not be able to participate in the experiment when circles in different colors are used. This problem does not occur when using different shapes to indicate the different sets of data in the separation monitor.

During the activation of the automation, all aircraft pairs on the separation monitor are interactive. This means that clicking on an aircraft pair in the separation monitor highlights both aircraft on the PVD. This helps the controller to understand what the automation is trying to achieve with the resolution advisory. Furthermore, this functionality should help create visual momentum for the controller, as they have to extract and integrate information across different displays and this creates a link between the displays. As the PVD contains colors to represent the controller solution and the automation solution, these same colors will be used for the separation monitor previews. Even though this will add nothing for participants suffering from colorblindness, it creates extra visual momentum for those who can distinguish the colors (Woods, 1984).

B-5 Projected aircraft path functionality

Using the adapted separation monitor increases the transparency of the AA tool, but it also provides the operator with data on the future state of the sector. As this information on the future state of the sector might have an effect on operator acceptance and frustration by itself, another experimental condition is proposed in which the separation monitor will not be shown.

Information on the future state of the sector is given by means of projected future aircraft positions. The projections assume a linear flight path with constant speed and are available for the current situation, the situation when the resolution advisory would be implemented and the situation when the operator's solution would be implemented. An example of these projected future aircraft positions can be seen in figure B-5. Projected positions are available for the same amount of time as available on the x-axis of the separation monitor. While the projected aircraft path functionality does give information on the future state of the sector, it does not give any information on the logic used by the AA algorithm.

This projected aircraft path functionality will always be available to the operator for the current traffic situation. When the automation becomes active, the projected aircraft path functionality will also be available for the automation solution and proposed operator solutions. Figure B-6 shows the projected aircraft path functionality when the automation is active.

In figure B-6, the magenta preview represents the projected path for the operator solution and the orange preview represents the projected path for the automation solution. The data that results from using these projections can then be used to determine the effect of information on the future state of the sector. This allows the effect of the increased transparency provided by the separation monitor to be determined independently of the effect of information on the future state of the sector.

Next to representing an alternative means to provide information on the future state of the sector, the projected aircraft path functionality can also be seen as a different means of transparency. It increases the sector transparency in the same way as the separation monitor does, namely by giving information on the future state of the sector, but it does not reflect the internal logic of the automation in any way. Using this functionality in an experiment can give information on whether or not it is necessary to show the internal logic of the automation

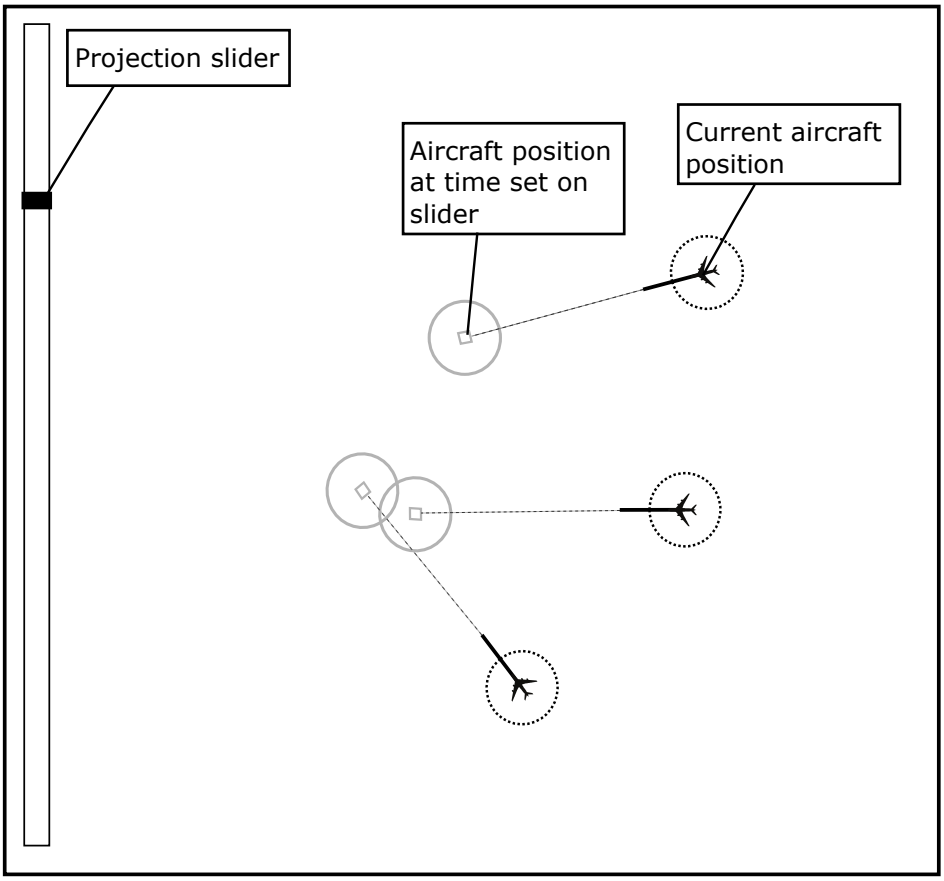


Figure B-5: Schematic of the projected future aircraft position functionality

when increasing transparency for this particular AA tool. The projected aircraft path function will use the same color scheme as the separation monitor.

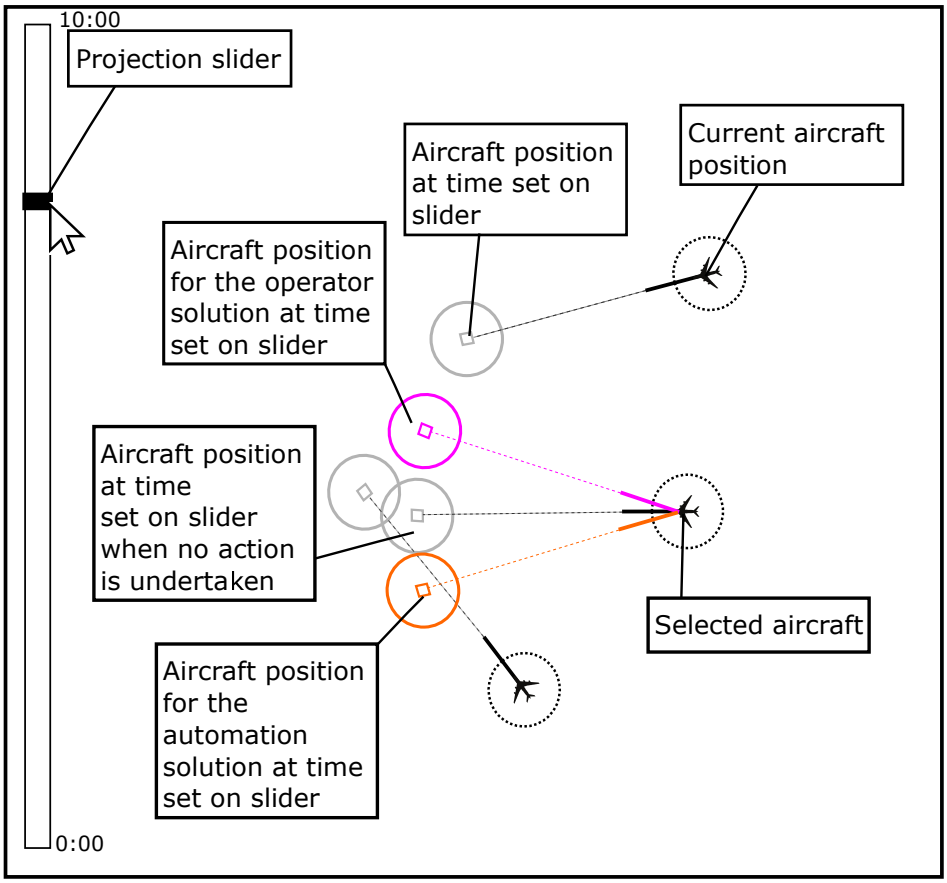


Figure B-6: Schematic of the projected future aircraft position functionality when the automation is active

Appendix C

Experiment design

This chapter will discuss the design of the experiment. The goal of this experiment is to determine the effect of automation transparency on operator acceptance, frustration and performance.

C-1 Experiment outline

In the experiment the participants will have two main tasks:

1. maintain safe separation between aircraft (5NM)
2. clear each aircraft to its respective sector exit point

To separate aircraft and steer them to their exit point, heading clearances can be given to aircraft. This can be done by clicking on an aircraft and dragging its velocity vector to a new location on a circular ring. To implement or activate a clearance, the ENTER key needs to be pressed on the keyboard. The exit points are shown as yellow triangles with four-letter names around the perimeter of the sector. The designated exit point of each aircraft is also shown in the flight label of the aircraft.

During the experiment, the automation aid assisted the operator in controlling the traffic. In one part of the experiment, the projected aircraft path functionality was available for use, as presented in section B-5. In the other part of the experiment, the adapted separation monitor was available for use. The adapted separation monitor will only be available when the automation is active. A basic separation monitor without interactive capabilities or the projected aircraft path functionality will be available to the participants at all times during the experiment. This decision has been made to ensure a fair comparison between the results from this experiment and those from the experiment of IJtsma (2016). It must however be noted that this may result in problems with SA in the case of the adapted separation monitor, as the participant suddenly has access to a lot of new information that they have to

quickly process in order to gain insight into the situation in this representation. This can be compared with a loss of SA, in terms of having to rebuild the mental picture of the situation. The availability of the basic separation monitor without interactive capabilities whenever the automation is not active should help to minimize these effects.

C-2 Research question and hypotheses

This research focuses on the following research question: What is the influence of transparency on the acceptance of and controller frustration with the AA tool developed by IJtsma (2016)? The following hypotheses will be tested in the experiment:

- Availability of the separation monitor results in higher operator acceptance of resolution advisories than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in lower operator frustration than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in lower operator workload than availability of the projected aircraft path functionality.
- Availability of the separation monitor results in higher operator trust in the automation than availability of the projected aircraft path functionality.
- Availability of the separation monitor or the projected aircraft path function results in higher operator acceptance of resolution advisories than when no automation transparency is available.
- Availability of the separation monitor or the projected aircraft path function results in lower operator frustration than when no transparency is available.
- Availability of the separation monitor or the projected aircraft path function results in lower operator workload than when no transparency is available.

C-3 Experiment variables

This section will list all relevant experiment variables. Subsection C-3-1 will describe the independent variables, subsection C-3-2 will list the dependent variables and subsection C-3-3 will list the control variables.

C-3-1 Independent variables

The experiment has two independent variables, namely the AA triggering threshold and the level of transparency. In the experiment of IJtsma (2016), the triggering threshold had four levels, namely a manual control condition and triggering of the AA after 1, 2 or 3 bad operator decisions. These levels were present in this experiment as well. This allows comparison of the effect of the 'preview' functionality and the basic separation monitor to the different

automation conditions. As this independent variable was a within-participants factor in the experiment of IJtsma (2016), this also was a within-participants factor in this experiment.

The level of transparency has two different levels, namely the preview functionality without the separation monitor and the adapted separation monitor. As IJtsma (2016) performed his experiment using a baseline level of transparency, it is not needed to include it in this experiment. As this experiment will be performed with different participants than the experiment of IJtsma (2016), this independent variable was a between-participants factor.

C-3-2 Dependent variables

The following dependent measures will be recorded:

- Interface
  - The total amount of time the projected aircraft path function is used.
  - The total number of times the projected aircraft path function is clicked on.
  - The total number of times aircraft pairs are clicked on in the separation monitor under active automation.
- Automation acceptance
  - CARS ratings will be asked for after each run **?**.
- Subjective workload
  - ISA ratings of workload will be asked for once every minute **?**.
  - NASA-TLX ratings will be asked for at the end of each experiment run **?**.
- Safety
  - The number and duration of STCAs
  - The number of duration of LoSs
  - The number of mid-air collisions (the number of times during a run when separation decreases below 0.5 NM)
- Automation
  - Automation operative time, which is the percentage of time that automation support is active.
  - The number of advisories, which can be subdivided into accepted, rejected and expired advisories.
- Airspace complexity
  - The relative no-go-area of the solution space can be used as a predictor of workload. This metric indicates how many vector commands in the available control space (defined as all possible vector commands within 360 degrees from the current heading) will result in a conflict with other aircraft. The metric considers the maximum no-go area of all aircraft in the airspace **?**.

- Situation awareness
  - An adaptation of the SASHA questionnaire will be completed after each run **?**.
- Trust
  - The Checklist for Trust between People and Automation developed by Jian et al. will be completed after each experiment run Jian et al. (2000).
- Frustration
  - A frustration questionnaire will be completed by participants after the experiment.

Previous research has indicated that people might unknowingly act 'politely' to a computer when the same computer asks for feedback on itself (Reeves & Nass, 1996). Therefore, the NASA-TLX ratings, Controller Acceptance Rating Scale (CARS) ratings and the checklist for trust will be completed on a paper form. It must also be noted that the NASA-TLX ratings include a measure for frustration. The checklist for trust has been included in appendix D.

C-3-3 Control variables

The following control variables will be kept constant:

- The automation will stay active for 30 seconds each time it is triggered. When the automation switches off while an advisory is still active, that advisory remains active until it either expires or is accepted or rejected by the subject. During this time, the preview on the separation monitor is available for these conflicts.
- Resolution advisories expire after 15 seconds.
- The additional margin that is used in constructing the resolution advisories is set at 2 NM.
- The traffic will be the same for each experiment condition, because this allows comparing the conditions without confound. To prevent the subjects from recognizing conflict geometries from earlier experiment runs, the airspace will be rotated (90, 180 or 270 degrees) between experiment runs. The traffic scenarios used will be the same traffic scenarios used in the experiment of IJtsma (2016), as this allows comparison between the results from this experiment and the experiment of IJtsma (2016) without confound.
- The basic separation monitor without preview functionality or interactive capabilities, or the basic projected aircraft path functionality will always be available to the controller, depending on the level of the independent variable transparency.
- The time scale on the x-axis of the separation monitor and the maximum time on the projected aircraft path slider are set to 10 minutes.

C-4 Experiment matrix

In table C-1, the experiment matrix can be seen. In this matrix, PAP stands for projected aircraft path, SM stands for separation monitor, TT stands for triggering threshold and M stands for manual control. The number at the end represents the degree of rotation of the sector. All experiment runs will be performed using the same sector and traffic geometry. However, the sector will be rotated for each consecutive run, preventing participant familiarization with the scenario. All conditions have been balanced among the participants.

Participant	Condition 1	Condition 2	Condition 3	Condition 4
1	PAP, TT1, 0	PAP, TT2, 90	PAP TT3, 180	PAP, M, 270
2	PAP, TT2, 180	PAP, TT1, 270	PAP, M, 0	PAP, TT3, 90
3	PAP, TT3, 0	PAP, M, 90	PAP, TT1, 180	PAP, TT2, 270
4	PAP, M, 180	PAP, TT3, 270	PAP, TT2, 0	PAP, TT1, 90
5	SM, TT1, 0	SM, TT2, 90	SM, TT3, 180	SM, M, 270
6	SM, TT2, 180	SM, TT1, 270	SM, M, 0	SM, TT3, 90
7	SM, TT3, 0	SM, M, 90	SM, TT1, 180	SM, TT2, 270
8	SM, M, 180	SM, TT3, 270	SM, TT2, 0	SM, TT1, 90

Table C-1: Experiment matrix

C-5 Participants

Sixteen participants (3 female, 13 male) were selected for this experiment, consisting of students and staff members of the Faculty of Aerospace Engineering. All subjects had some experience with directing air traffic in an ATC simulation, either through participating in courses or earlier experiments related to ATC.

C-6 Participant briefing and training

Before the experiment, the participants were briefed on the ATC simulator and how to use it. The information given to the participant was roughly the same as in the pre-experiment briefing of IJtsma (2016), with the addition of information on the projected aircraft path functionality or separation monitor, depending on the experiment group. The training the participants received has also been roughly the same as the training the participants of the experiment of IJtsma (2016) received.

Appendix D

Checklist for Trust between People and Automation

The study of Jian et al. (2000) consisted of three parts. In the first part, participants were asked for written descriptions of their notion of trust and distrust. The participants in this study were graduate students majoring in linguistics or English, because of their knowledge of word meanings. Furthermore, these same participants were asked to rate a set of 138 words on whether or not they thought these words related to trust in either a positive or negative way. From this study, a set of 96 words related to trust was identified.

The second part of the study was conducted using 120 participants from the university community. They were asked to which extent they thought the words in the set identified in phase one were related to trust or distrust. It was found that trust and distrust were highly negatively correlated, indicating that trust and distrust are opposites, instead of different factors. Furthermore, there was a difference between human-human trust and human-machine trust.

The third part of the study was conducted using 30 participants from the university community. These participants were asked to compare the similarity of 30 words that were related to trust and distrust most strongly from the results of part two of the study. A factor analysis identified eight significant words related to human-machine trust. Furthermore, a cluster analysis was used to group similar words. Based on this cluster analysis, a trust scale including 12 items was proposed. The resulting checklist for trust between people and automation can be seen in figure D-1.

All participants in the experiments of Jian et al. (2000) were native English speakers. As participants in the experiment proposed in this report will be sought at the faculty of aerospace engineering at the Delft University of Technology, most of the participants will not be native English speakers. As the majority of participants will most likely be native Dutch speakers, a Dutch translation of this checklist for trust is needed to ensure proper understanding of the questions.

In order to obtain the best translation possible with the limited resources available, the questions in this checklist will not be directly translated. Even translating the 30 words used

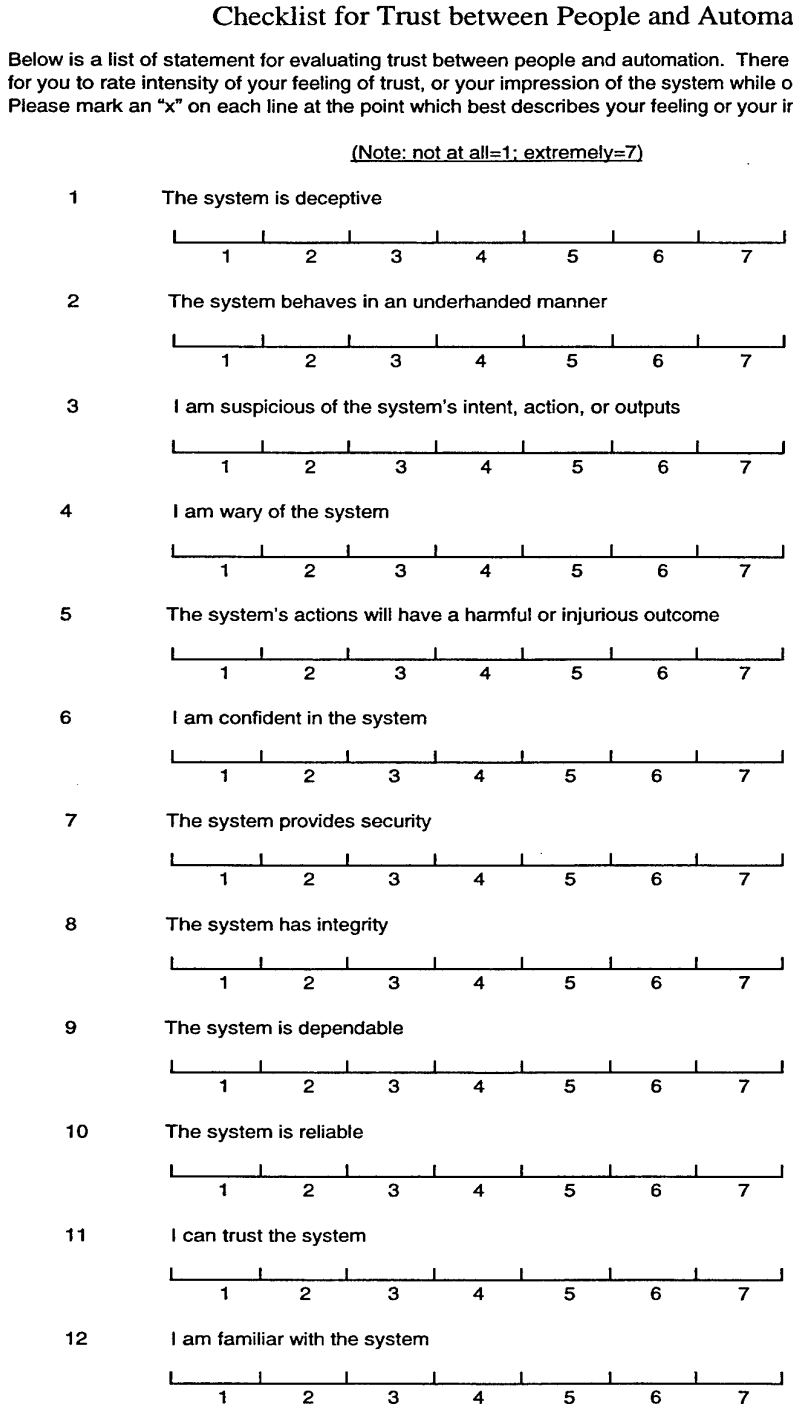


Figure D-1: The checklist for trust between people and automation (Jian et al., 2000)

in part 3 of the study and repeating this part of the study is not feasible with the resources

available. Therefore, the words in the word clusters identified with the cluster analysis in part 3 of the study will be translated and the same word clusters will be used to create the Dutch version of the checklist for trust. The resulting Dutch version of the checklist for trust will therefore not be a direct translation of the English version. Table D-1 shows table 5 from the paper of Jian et al. (2000).

The right column of table D-1 shows the clusters of words that were identified in the cluster analysis. The left column holds the questions related to each cluster of words. The right column of table D-2 holds the Dutch translations of the words in the right column of table D-1. The left column of table D-2 shows the questions for the Dutch checklist for trust, which are based on the terms in the right column of this table and the questions in the left column of table D-1.

Item	Word groups from cluster analysis
The system is deceptive.	Deception Lie Falsity Betray Misleading Phony Cheat
The system behaves in an underhanded manner.	Sneaky Steal
I am suspicious of the system's intent, action, or output.	Mistrust Suspicion Distrust
I am wary of the system.	Beware
The system's action will have a harmful or injurious outcome.	Cruel Harm
I am confident in the system.	Assurance Confidence
The system provides security.	Security
The system has integrity.	Honor Integrity
The system is dependable.	Fidelity Loyalty
The system is reliable.	Honesty Promise Reliability Trustworthy Friendship Love
I can trust the system.	Entrust
I am familiar with the system.	Familiarity

Table D-1: Trust scale items for human-machine trust and the corresponding cluster of trust related words on which they were based (Jian et al., 2000)

Item	Word groups from cluster analysis
Het systeem bedriegt mij.	Bedrog Leugen Incorrectheid Verraad Misleidend Nep Valsspelen
Het systeem voert ongezien acties uit.	Stiekem Stelen
Ik vind de bedoelingen, acties, of uitvoer van het systeem verdacht.	Mistrouwen Verdacht Wantrouwen
Ik ben op mijn hoede als ik met het systeem werk.	Waakzaamheid
De acties van het systeem hebben een gevaarlijke of schadelijke uitkomst.	Gemeen Gevaar
Het systeem ondersteunt mij.	Zekerheid Vertrouwen
Het systeem zorgt voor een gevoel van veiligheid.	Veiligheid
Het systeem is integer.	Eer Integriteit
Het systeem is betrouwbaar.	Trouw Loyaliteit
Ik vertrouw het systeem.	Eerlijkheid Belofte Deugdelijk Vertrouwenswaardig Vriendschap Liefde
Ik durf taken aan het systeem toe te vertrouwen.	Toevertrouwen
Ik ben vertrouwd met het systeem.	Vertrouwdheid

Table D-2: Dutch translation of the clusters of words and the items for the checklist for trust

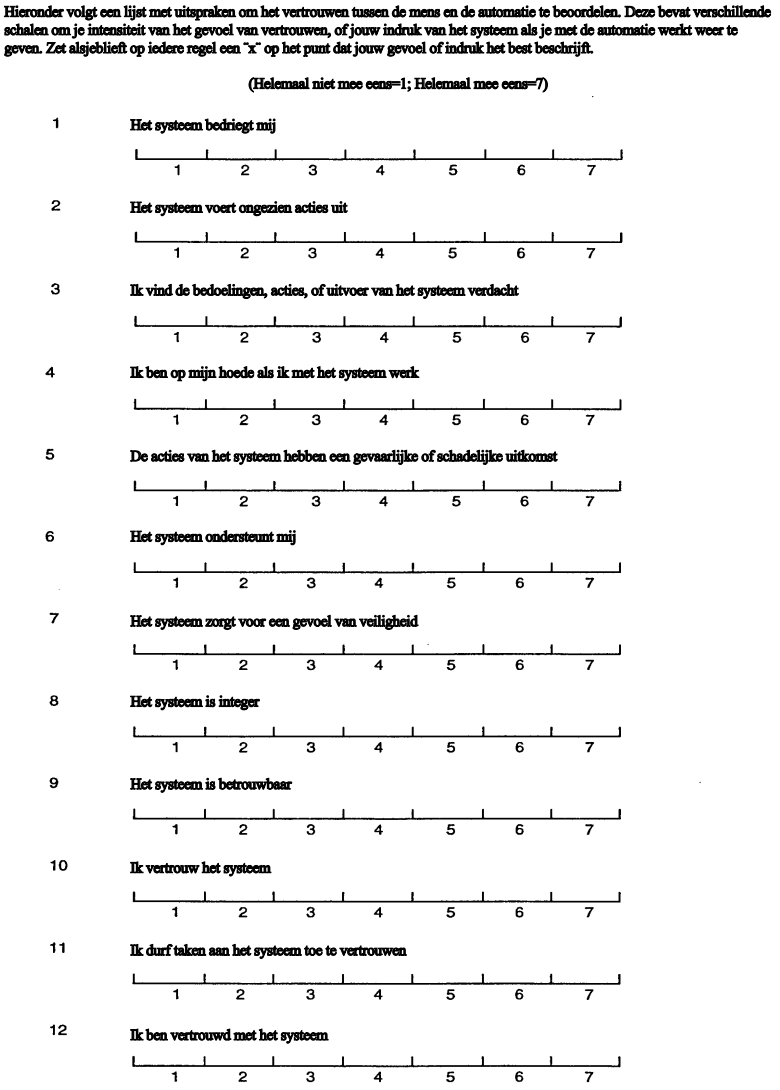


Figure D-2: The Dutch translation of the checklist for trust between people and automation

Appendix E

Pre-experiment briefing

As this experiment has a between-participants factor, a different pre-experiment briefing has been created for each of the conditions. The first pre-experiment briefing shown is the one that participants using the separation monitor have received, while the second pre-experiment briefing is the one that participants using the projected aircraft path functionality have received.

Pre-experiment briefing

Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control

This experiment is meant to evaluate the effect of a designed automation algorithm for Air Traffic Control (ATC) on the workload, safety and performance of air traffic controllers. This briefing contains a short overview of the experiment, that is, some explanation about the control task, the automation support and the experiment procedure.

Apparatus

The experiment will be conducted in the Air Traffic Management Laboratory (ATMLab), on the second floor of the SIMONA-building. This lab uses LCD screens on which a simulated ATC radar screen is shown. As a subject, you can interactively control aircraft trajectories using the mouse and keyboard. An example of the simulated radar screen is shown in figure 1.

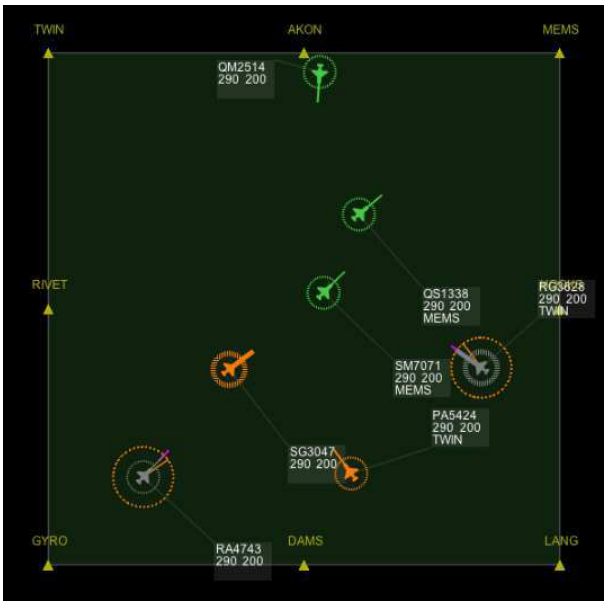


Figure 1: Example of simulated radar screen.

Control task

During the experiment you will act as an air traffic controller. You will thus control the trajectories of the various aircraft in your sector. You will have two responsibilities:

1. Maintain a separation of at least 5 NM between aircraft at all times (the minimum required separation).
2. Guide all aircraft to their respective exit waypoints in the sector, as efficiently as possible.

To fulfill these responsibilities, you will only be allowed to make heading changes to aircraft (so no altitude and speed changes). These heading changes can be given to aircraft by clicking on an aircraft and subsequently dragging its speed vector towards the desired heading. When you press the ENTER key, the heading change is send to the aircraft and implemented immediately (corresponding to a situation in which ADS-B is used to communicate and implement vector commands).

Automation Support

During the experiment, an automation aid will assist you in controlling the traffic. This is a so-called adaptive automation algorithm, that provides you with support based on your performance. Performance is assessed by looking at how well you resolve conflicts between aircraft, focusing on whether you introduce additional conflicts in the sector as a consequence of your actions. For instance, if you vector an aircraft in a certain direction but as a consequence of this vectoring two new conflicts are created, your action has a nett adverse effect. This is a trigger to provide you with automation support. Once activated, the automation support will be active for 30 seconds.

Separation monitor

During the experiment, the separation monitor is available to you. The separation monitor can be seen in figure 2.

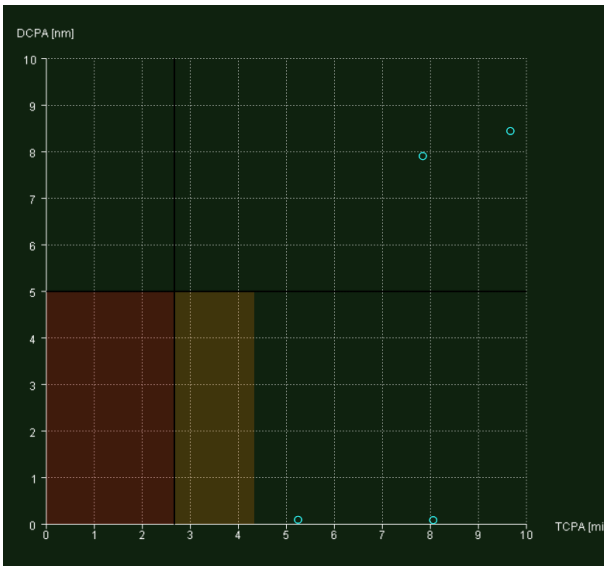


Figure 2: The separation monitor.

In figure 2, each circle represents a pair of aircraft. The position of the circle on the y-axis indicates the distance between the two aircraft at their closest point of approach (CPA). When an aircraft pair appears below the dashed line, it will violate the separation minimum at some time in the future. The position of the circle on the x-axis indicates how long it will take for the two aircraft to reach their CPA. However, when

the aircraft pair will violate the separation minimum at some point in the future, the time indicated is the time until a loss of separation occurs. The red area indicates the occurrence of a short-term collision alert for a particular aircraft pair and the orange area indicates a medium-term collision alert.

When the automation support becomes active, a number of changes occur on the separation monitor. Figure 3 shows the separation monitor when the automation support is active.

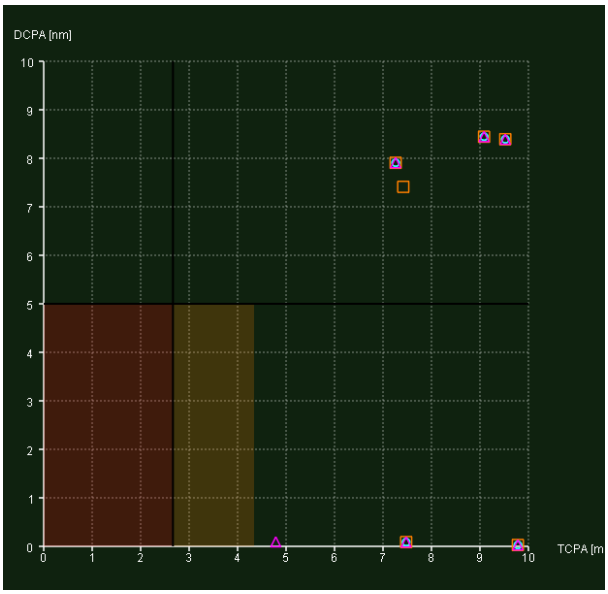


Figure 3: The separation monitor when the automation support is active.

The orange squares on the separation monitor indicate the traffic situation that would occur when the solution of the automation support is implemented. The magenta triangles on the separation monitor indicate the traffic situation that would occur when the solution you proposed would be implemented. The circles still show the current traffic situation. Furthermore, whenever the automation support is active, you can click on any symbol in the separation monitor. While you are clicking a symbol, the two aircraft that pertain to that symbol will be highlighted on the radar screen. When the automation support deactivates, these extra functionalities will be deactivated as well.

Experiment Conditions

A total of four experiment conditions will be tested during the experiment. The conditions offer the same type of automation support, but differ in when the support is provided. Each condition will be tested once and will be presented to you in a random order.

Experiment Procedure

The experiment will start with a short briefing about the air traffic control task. Some best practices will be explained, along with the rationale behind the automation support. After the briefing there will be

a training phase. You will be allowed to familiarize yourself with the control task, the best practices, the automation support and the separation monitor. This will also give you the chance to try out different control strategies and see which control strategies work best for you. During the training phase, the complexity of the scenarios will gradually increase. After the training phase, a short break with be held. After the break, the measurement runs will start. There will be four measurement runs, one for each experiment condition, in a random order. During the measurement runs, a pop-up window will frequently appear asking you to fill out the workload you are currently experiencing. Additionally, after each experiment run, you will be asked to fill out three forms:

1. NASA Task Load Index (NASA-TLX) form, to get an indication of the general workload you experienced during that run.
2. SASHA questionnaire, to get an indication of your situation awareness during the run.
3. Controller Acceptance Rating Scale (CARS), which asks you to rate the automation support.
4. Checklist for Trust between People and Automation, to asses your trust in the automation.

It is very important that you carefully fill out these forms. In the appendix of this briefing you will find an example of each of these forms. The experiment will be completed in approximately three hours, including breaks. If during the experiment there is any instance you feel uncomfortable for whatever reason, you can stop at any time.

Thanks a lot for your participation and good luck with the experiment!

Pre-experiment briefing

Effect of Automation Transparency on Acceptance of Resolution Advisories in Air Traffic Control

This experiment is meant to evaluate the effect of a designed automation algorithm for Air Traffic Control (ATC) on the workload, safety and performance of air traffic controllers. This briefing contains a short overview of the experiment, that is, some explanation about the control task, the automation support and the experiment procedure.

Apparatus

The experiment will be conducted in the Air Traffic Management Laboratory (ATMLab), on the second floor of the SIMONA-building. This lab uses LCD screens on which a simulated ATC radar screen is shown. As a subject, you can interactively control aircraft trajectories using the mouse and keyboard. An example of the simulated radar screen is shown in figure 1.

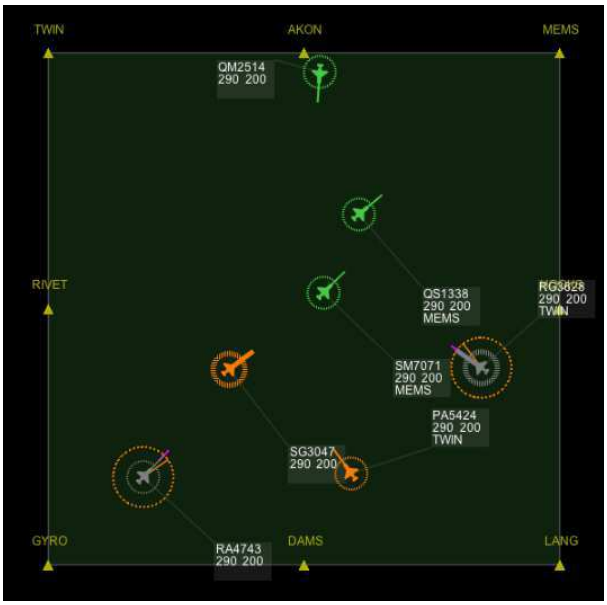


Figure 1: Example of simulated radar screen.

Control task

During the experiment you will act as an air traffic controller. You will thus control the trajectories of the various aircraft in your sector. You will have two responsibilities:

1. Maintain a separation of at least 5 NM between aircraft at all times (the minimum required separation).
2. Guide all aircraft to their respective exit waypoints in the sector, as efficiently as possible.

To fulfill these responsibilities, you will only be allowed to make heading changes to aircraft (so no altitude and speed changes). These heading changes can be given to aircraft by clicking on an aircraft and subsequently dragging its speed vector towards the desired heading. When you press the ENTER key, the heading change is send to the aircraft and implemented immediately (corresponding to a situation in which ADS-B is used to communicate and implement vector commands).

Automation Support

During the experiment, an automation aid will assist you in controlling the traffic. This is a so-called adaptive automation algorithm, that provides you with support based on your performance. Performance is assessed by looking at how well you resolve conflicts between aircraft, focusing on whether you introduce additional conflicts in the sector as a consequence of your actions. For instance, if you vector an aircraft in a certain direction but as a consequence of this vectoring two new conflicts are created, your action has a nett adverse effect. This is a trigger to provide you with automation support. Once activated, the automation support will be active for 30 seconds.

Projected aircraft path function

During the experiment, the projected aircraft path function is available to you. The projected aircraft path function can be seen in figure 2.

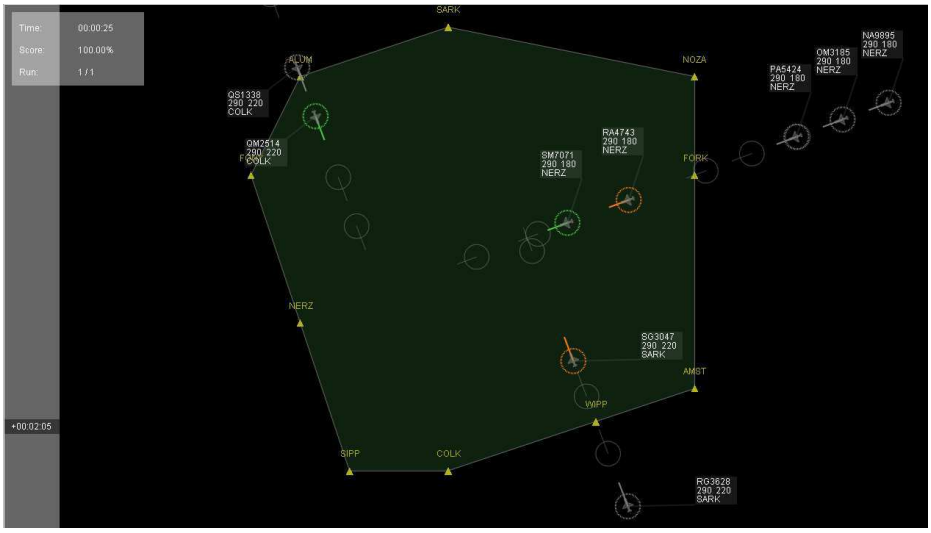


Figure 2: The projected aircraft path function.

In figure 2, the slider bar on the left controls the projected aircraft path function. When this bar is dragged upwards, a projection of the positions of all aircraft on the radar screen will be shown. As the bar is dragged higher up, the projection will show the positions of all aircraft further into the future. In figure 2 it can be seen that the two aircraft highlight in orange will experience a loss of separation in about 2 minutes.

When the automation support becomes active, The projected aircraft path function changes. Figure 3 shows the projected aircraft path function when the automation support is active.

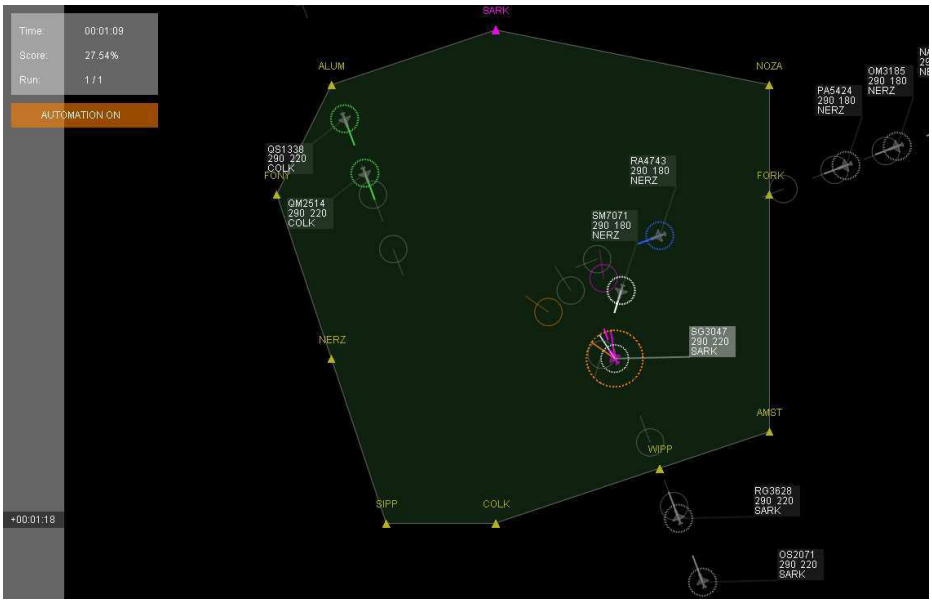


Figure 3: The projected aircraft path function when the automation support is active.

The projected aircraft path now shows three different colored aircraft projections for aircraft that have received advisories and/or you have selected. The gray projection shows the aircraft position when the current heading is maintained. The orange projection shows the aircraft position when the automation advisory is implemented and the magenta projection shows the aircraft position when the solution you proposed would be implemented. These extra projections are only available whenever the automation support is active, so they disappear when the automation support deactivates.

Experiment Conditions

A total of four experiment conditions will be tested during the experiment. The conditions offer the same type of automation support, but differ in when the support is provided. Each condition will be tested once and will be presented to you in a random order.

Experiment Procedure

The experiment will start with a short briefing about the air traffic control task. Some best practices will be explained, along with the rationale behind the automation support. After the briefing there will be a training phase. You will be allowed to familiarize yourself with the control task, the best practices, the automation support and the separation monitor. This will also give you the chance to try out different control strategies and see which control strategies work best for you. During the training phase, the complexity of the scenarios will gradually increase. After the training phase, a short break with be held. After the break, the measurement runs will start. There will be four measurement runs, one for each experiment condition, in a random order. During the measurement runs, a pop-up window will frequently appear asking you to fill

out the workload you are currently experiencing. Additionally, after each experiment run, you will be asked to fill out three forms:

1. NASA Task Load Index (NASA-TLX) form, to get an indication of the general workload you experienced during that run.
2. SASHA questionnaire, to get an indication of your situation awareness during the run.
3. Controller Acceptance Rating Scale (CARS), which asks you to rate the automation support.
4. Checklist for Trust between People and Automation, to asses your trust in the automation.

It is very important that you carefully fill out these forms. In the appendix of this briefing you will find an example of each of these forms. The experiment will be completed in approximately three hours, including breaks. If during the experiment there is any instance you feel uncomfortable for whatever reason, you can stop at any time.

Thanks a lot for your participation and good luck with the experiment!

Appendix F

Additional experiment results

This chapter will show some additional results from the experiment that were not presented in the paper. Analysis for these parameters has only been performed for the separation monitor and projected aircraft path groups.

F-1 Average extra pathlength

Figure F-1 shows boxplots for the average additional pathlength for all aircraft. The additional path length is defined as the extra distance an aircraft flew when comparing its path to a straight line from entry to exit waypoint. It can be seen that for the separation monitor group, additional pathlength appears to decrease slightly for the stricter automation conditions. For the projected aircraft path group, additional pathlength appears to be lower for the strictest automation condition, but additional pathlength appears to be higher for the other automation conditions when compared to the manual control condition. Overall, additional pathlength appears to be slightly lower for the projected aircraft path group than for the separation monitor group.

AFriedman's two-way ANOVA shows that there is no significant difference in average additional pathlength between the different automation conditions. Furthermore, a Kruskal-Wallis test shows no significant difference in average additional pathlength between the separation monitor and projected aircraft path groups for any of the automation conditions.

F-2 Variance of Z-scored ISA ratings

Figure F-2 shows boxplots for the variance of the z-scored ISA workload ratings. It can be seen that, for the separation monitor group, variance in ISA z-scores appears to decrease slightly with increasing automation strictness. However, for the projected aircraft path group, variance in z-scored ISA workload ratings appears to increase slightly for the stricter automation conditions.

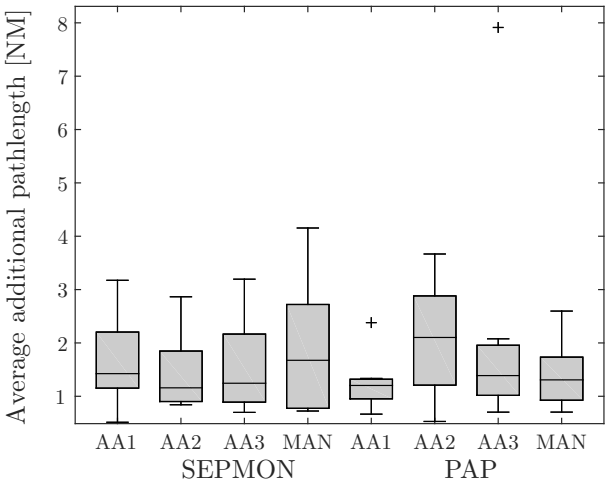


Figure F-1: Boxplots for the average additional pathlength

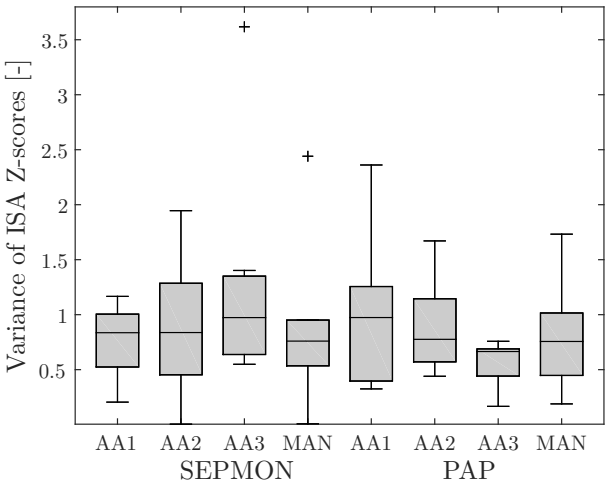


Figure F-2: Boxplots for the variance of the z-scored ISA workload ratings

A Friedman's two-way ANOVA shows that there is no significant difference in variance of z-scored ISA workload ratings between the different automation conditions. Furthermore, a Kruskal-Wallis test shows no significant difference in variance of z-scored ISA workload ratings between the separation monitor and projected aircraft path groups for any of the automation conditions.

F-3 Average maximum distance between aircraft

Figure F-3 shows boxplots for the average maximum distance between aircraft. Overall, it appears that the average maximum distance between aircraft increases for the stricter automation conditions. Furthermore, average maximum distance between aircraft appears to be slightly higher for the separation monitor group than for the projected aircraft path group.

A repeated measures ANOVA shows that there is no significant difference in average maximum distance between aircraft for the different automation conditions. Furthermore, no significant

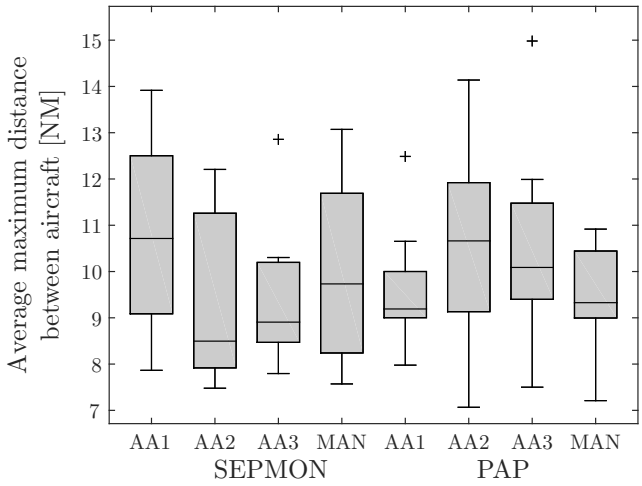


Figure F-3: Boxplots for the average maximum distance between aircraft

difference in average maximum distance between aircraft was found between the separation monitor and projected aircraft path groups.

F-4 Number of LOA changes

Figure F-4 shows boxplots of the number of LOA changes. A clear trend of more LOA changes for the stricter automation conditions can be seen. Furthermore, it appears that the projected aircraft path group experienced fewer LOA changes than the separation monitor group.

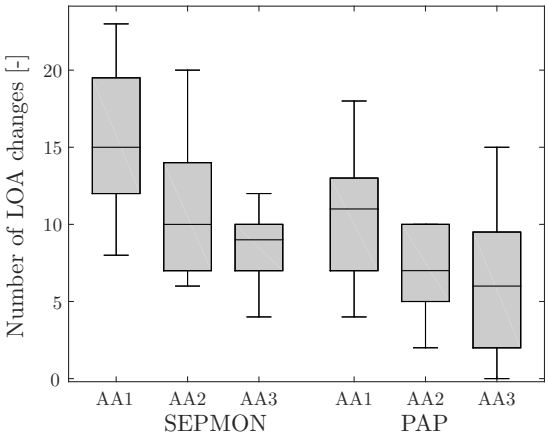


Figure F-4: Boxplots for the number of LOA changes

A repeated measures ANOVA shows that there is indeed a significant difference in number of LOA changes between the different automation conditions ($F(2)=9.255$; $p<0.05$). Furthermore, a significant difference in number of LOA changes exists between the separation monitor and projected aircraft path groups ($F(1)=5.909$; $p<0.05$).

F-5 Total number of commands

Figure F-5 shows boxplots of the total number of commands given by the participants. There appears to be a slight decrease in number of commands given for the stricter automation conditions. Furthermore, the number of commands given appears to be lower for the projected aircraft path group than for the separation monitor group.

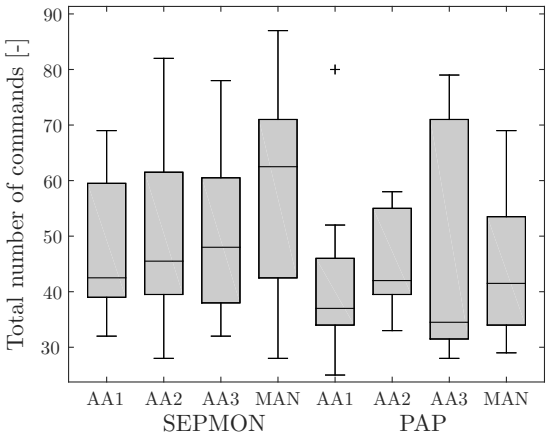


Figure F-5: Boxplots for the total number of commands

A repeated measures ANOVA shows that there is no significant difference in total number of commands given between the different automation conditions. Furthermore, no significant difference in total number of commands given exists between the separation monitor and projected aircraft path groups.

Appendix G

Questionnaire results

This appendix holds some additional results for the NASA-TLX, adapted SASHA questionnaire, the checklist for trust and the frustration questionnaire.

G-1 NASA-TLX

In Figures 17, 18 and 19, it appears that a correlation between the weights and the ratings for the different sources of load on the NASA-TLX questionnaire is present. A correlation analysis on these factors has thus been performed. Table G-1 shows the results of the correlation analysis.

Source of load	Pearson correlation	p-value
Mental demand	0.093	0.369
Physical demand	0.718	0.000
Temporal demand	0.451	0.000
Performance	-0.141	0.171
Effort	0.466	0.000
Frustration	0.485	0.000

Table G-1: Pearson correlations for weights and ratings for the NASA-TLX components

It can be seen that for four out of the six sources of load, a correlation exists between the weight and rating given in the NASA-TLX questionnaire.

G-2 SASHA

This section holds additional results for the adapted SASHA questionnaire. Subsection G-2-1 hold the statistical results for the adapted SASHA questionnaire and subsection G-2-2 holds the comments participants gave to the questions.

G-2-1 Statistical results

Table G-2 shows the results of the Friedman's two-way ANOVA for the questions on the adapted SASHA questionnaire. It can be seen that there is no significant difference between the different automation conditions for any of the questions.

Question number	Test statistic	p-value
1	$\chi^2(3)=0.080$	0.994
2	$\chi^2(3)=1.812$	0.612
3	$\chi^2(3)=1.054$	0.788
4	$\chi^2(3)=4.979$	0.173
5	$\chi^2(3)=0.117$	0.990
6	$\chi^2(3)=1.549$	0.671
7	$\chi^2(3)=1.783$	0.619
8	$\chi^2(3)=3.667$	0.300
9	$\chi^2(3)=0.857$	0.836
10	$\chi^2(3)=3.191$	0.363
11	$\chi^2(3)=0.193$	0.979
12	$\chi^2(3)=3.028$	0.387

Table G-2: Friedman's two-way ANOVA results for the adapted SASHA questionnaire

Table G-3 shows the results of the Kruskal-Wallis test for the questions on the adapted SASHA questionnaire. It can be seen that there is no significant difference between the different experiment groups for any of the questions for any of the automation conditions, except one. A significant difference in answers to question 10 for condition AA2 was found.

G-2-2 Comments

For each question on the adapted SASHA questionnaire, participants had the option to add a comment. These comments are listed below.

Q1: P14: By the end it got more messy, so it became harder to take time to predict traffic.

Q2: P3: I was able to organize the traffic the way I wanted by rejecting the AA suggestions once. P8: In a busy situation, the automation didn't really help. P9: Would be nice to always be able to see the pairs. P21: Multiple conflicts made it quite cumbersome at some point.

Q3: P3: At certain point I was also relying on automation to support in conflict detection. P6: just once. more than one time! a few times, I would say 2-3. P10: The conflict where there was a loss of separation. P14: When solving self-induced conflicts; Few times in a chain reaction of solution $->$ conflict etc. P21: The automation was indicating conflicts with targets outside my airspace - good but unexpected.

Q4: P8: Sometimes a three- or four-way conflict demanded all my attention. P21: When vectoring aircraft in acute cross angles.

Q5: P4: sometimes, after aircraft has been already vectored to its exit point, it stops being vectored to this exit point (shifts from given direction). P8: Twice, once very late; A lot,

Question number	Automation condition	Test statistic	p-value
1	AA1	H(1)=0.089	0.765
	AA2	H(1)=1.929	0.165
	AA3	H(1)=0.551	0.458
	MAN	H(1)=0.111	0.739
2	AA1	H(1)=0.937	0.333
	AA2	H(1)=1.967	0.161
	AA3	H(1)=0.922	0.337
	MAN	H(1)=1.478	0.224
3	AA1	H(1)=0.012	0.912
	AA2	H(1)=0.300	0.584
	AA3	H(1)=0.027	0.870
	MAN	H(1)=0.505	0.477
4	AA1	H(1)=0.012	0.912
	AA2	H(1)=0.362	0.547
	AA3	H(1)=0.015	0.903
	MAN	H(1)=0.156	0.693
5	AA1	H(1)=2.552	0.110
	AA2	H(1)=0.059	0.808
	AA3	H(1)=0.000	1.000
	MAN	H(1)=0.682	0.409
6	AA1	H(1)=0.529	0.467
	AA2	H(1)=2.049	0.152
	AA3	H(1)=0.059	0.809
	MAN	H(1)=0.318	0.573
7	AA1	H(1)=0.014	0.907
	AA2	H(1)=1.222	0.269
	AA3	H(1)=0.116	0.733
	MAN	H(1)=0.445	0.505
8	AA1	H(1)=2.679	0.102
	AA2	H(1)=3.857	0.050
	AA3	H(1)=0.003	0.955
	MAN	H(1)=1.929	0.165
9	AA1	H(1)=0.056	0.813
	AA2	H(1)=0.059	0.807
	AA3	H(1)=1.638	0.201
	MAN	H(1)=0.439	0.508
10	AA1	H(1)=0.248	0.618
	AA2	H(1)=3.990	0.046
	AA3	H(1)=1.190	0.275
	MAN	H(1)=0.153	0.696
11	AA1	H(1)=0.588	0.443
	AA2	H(1)=2.959	0.085
	AA3	H(1)=0.199	0.656
	MAN	H(1)=0.077	0.781
12	AA1	H(1)=0.942	0.332
	AA2	H(1)=1.621	0.203
	AA3	H(1)=1.127	0.288
	MAN	H(1)=0.000	1.000

Table G-3: Kruskal-Wallis test results for the adapted SASHA questionnaire

three times at least; Once, a little late; See above. P10: I responded late after the conflict resolution. P21: Yes, this time completely.

Q6: P6: never got it I think??? P8: Mostly, but not always. P10: Dint need the support; Dint need the support. P13: when an aircraft was close to the border but not in the way of it's waypoint the automation gave an advisory towards the sector border which would lead the aircraft to go outside the sector before the conflict was solved, since you first have to accept or reject the automation, this leaded to a higher workload since i tried to solve a conflict quickly. P14: Rarely activated during this run; The one time the automation provided a suggestion, I did not see why it was better than my current path (as I felt my current plan was more future-proof in terms of future collision prevention, and the plan of the automation

was not). P21: At some point I was considering a later conflict but had to spend some time to understand what the automation was trying to convey - was useful but had to think more about tit; Again the look-ahead time(just once) provided a resolution geometry I did not see fit but due to other traffic I had to attend to I accepted. Hence I vectored that aircraft once more a few minutes later.

Q7: P6: n/a. P8: Slight strategy change went a long way; I got what was going on, but I didn't necessarily agree with the solution. P10: Dint need the support; Dint need the support. P14: Sometimes, as in certain cases I felt that my control input would have accomplished the same. P21: Short term , yes.

Q8: P6: never felt a need for automation to pop up and it did not, so. P8: Wasn't there; Could've used it before I'd change a heading back to a waypoint only to find myself in a red conflict; The help was in time. P9: Because sometime I did not know which aircraft belonged to certain dots on the separated view; Sometimes needed to check the pairs. P10: Sometimes, I felt the AA interrupted when I was planning a resolution looking ahead of time; Dint need the support; Dint need the support, meaning it recognized when not to intervene. P14: I now know roughly when I might expect it to activate (which is when I've made a resolution to solve a conflict, and I want to reroute the aircraft back to its destination exit point -¿ in doing this, I might induce a conflict that I had not foreseen); In one scenario where my input led to a 5 second loss of separation warning, the automation did not take over; Hardy needed during this run.

Q9: P4: when projection of aircraft path is on, it is hard to distinguish which aircraft is were (each aircraft projection). P6: n/a. P8: It confused me once. P14: I would mostly hope the automation is well able to help me solve the described conflict above, I already feel I have a good understanding of these situations.

Q10: P6: did not trigger and I could see why, because of time slider mostly. P8: Not applicable, I'd say. P9: Sometimes in a later stage able to determine which pair of aircraft are approaching the orange zone. P10: Dint need the support; Dint need the support. P14: The aircraft in blue clarified a lot; Not as much the interface as my own expectation; If it triggers yes, but I don't know why it did not trigger; colorization helped.

Q11: P6: could understand using time slider that automation did not pop up. P8: Sometimes suggested solutions were radically different from the solution I had in mind, and accepting a solution would only make things worse (as in the amount of re-routing to be done, not in terms of safety) at the expense of a short-term solution; I understood what it wanted to do, but not always why it wanted to do that. P10: Dint need the support; Dint need the support. P14: Like I said, I found that sometimes my control action would have accomplished the same; See earlier (longer) comment.

Q12: P3: I was not satisfied with my SA. I had the feeling that I needed to update my plan according the automation's decisions, so I lost some focus a couple of times. P6: not 100%. P8: The automation was helpful for most of the simulation, but from 12:00 to 14:00 ish, it compounded the issues. P21: 3,5.

G-3 Checklist for trust between people and automation

Table G-4 shows the results of the Friedman's two-way ANOVA for the questions on the checklist for trust. It can be seen that there is no significant difference between the different automation conditions for any of the questions.

Question number	Test statistic	p-value
1	$\chi^2(3)=6.111$	0.106
2	$\chi^2(3)=4.219$	0.239
3	$\chi^2(3)=0.360$	0.948
4	$\chi^2(3)=4.810$	0.186
5	$\chi^2(3)=0.145$	0.986
6	$\chi^2(3)=0.872$	0.832
7	$\chi^2(3)=1.622$	0.654
8	$\chi^2(3)=2.182$	0.536
9	$\chi^2(3)=3.228$	0.358
10	$\chi^2(3)=0.877$	0.831
11	$\chi^2(3)=1.905$	0.592
12	$\chi^2(3)=7.596$	0.055

Table G-4: Friedman's two-way ANOVA results for the checklist for trust

Table G-5 shows the results of the Kruskal-Wallis test for the questions on the checklist for trust for the different displays. It can be seen that there is no significant difference between the different experiment groups for any of the questions for any of the automation conditions.

Figure G-1 shows the results of the checklist for trust for the different language groups. The gray boxplots show the results for the group that filled out the checklist in English, while the white boxplots show the results for the group that filled out the checklist in Dutch.

It can be seen in figure G-1 that the answers for the two language groups are quite similar for the different questions, with the exception of question four. It appears that the group filling out the Dutch version of the checklist give a higher score for this question than the group filling out the English version.

Table G-6 shows the results of the Kruskal-Wallis test for the questions on the checklist for trust for the different languages. It can be seen that there is no significant difference between the different experiment groups for any of the questions for any of the automation conditions, with the exception of question four.

Results show that participants in filling out the checklist for trust in Dutch gave significantly higher answers to question four than the participants filling out the checklist for trust in English. This might mean that the translation of question four is not accurate enough and is interpreted differently than the question in English. Overall, it appears that the translation of the checklist for trust is good, as no significant difference in answers was found for any of the other questions.

Question number	Automation condition	Test statistic	p-value
1	AA1	H(1)=0.000	1.000
	AA2	H(1)=0.000	1.000
	AA3	H(1)=0.000	1.000
	MAN	H(1)=0.385	0.535
2	AA1	H(1)=0.000	1.000
	AA2	H(1)=1.000	0.317
	AA3	H(1)=0.595	0.440
	MAN	H(1)=0.482	0.487
3	AA1	H(1)=0.432	0.511
	AA2	H(1)=0.000	1.000
	AA3	H(1)=0.238	0.626
	MAN	H(1)=0.835	0.361
4	AA1	H(1)=1.538	0.215
	AA2	H(1)=0.435	0.510
	AA3	H(1)=1.306	0.253
	MAN	H(1)=1.296	0.255
5	AA1	H(1)=0.170	0.680
	AA2	H(1)=0.056	0.814
	AA3	H(1)=1.485	0.223
	MAN	H(1)=0.170	0.680
6	AA1	H(1)=0.874	0.350
	AA2	H(1)=0.197	0.657
	AA3	H(1)=0.773	0.379
	MAN	H(1)=0.302	0.582
7	AA1	H(1)=0.003	0.955
	AA2	H(1)=0.245	0.621
	AA3	H(1)=1.921	0.166
	MAN	H(1)=0.542	0.462
8	AA1	H(1)=1.359	0.244
	AA2	H(1)=2.092	0.148
	AA3	H(1)=0.000	1.000
	MAN	H(1)=0.199	0.656
9	AA1	H(1)=0.319	0.572
	AA2	H(1)=0.112	0.738
	AA3	H(1)=1.490	0.222
	MAN	H(1)=0.004	0.952
10	AA1	H(1)=1.250	0.264
	AA2	H(1)=0.077	0.781
	AA3	H(1)=0.015	0.904
	MAN	H(1)=0.077	0.782
11	AA1	H(1)=0.054	0.816
	AA2	H(1)=0.885	0.347
	AA3	H(1)=0.148	0.700
	MAN	H(1)=0.243	0.622
12	AA1	H(1)=0.725	0.395
	AA2	H(1)=0.124	0.725
	AA3	H(1)=0.000	1.000
	MAN	H(1)=0.029	0.864

Table G-5: Kruskal-Wallis test results for the checklist for trust for the display condition

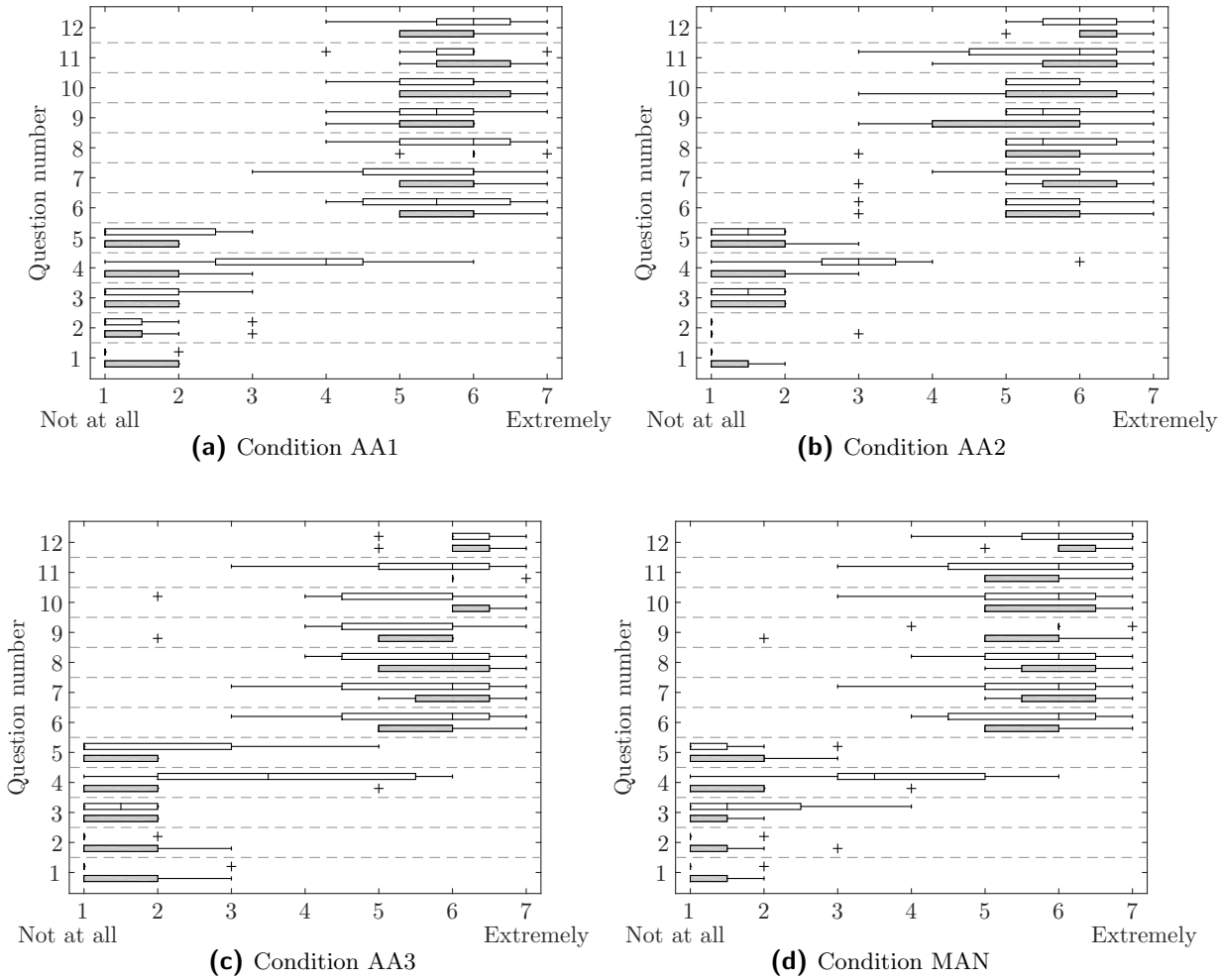


Figure G-1: Checklist for trust results for the different automation conditions

Question number	Automation condition	Test statistic	p-value
1	AA1	H(1)=1.250	0.264
	AA2	H(1)=2.143	0.143
	AA3	H(1)=0.937	0.333
	MAN	H(1)=0.385	0.535
2	AA1	H(1)=0.000	1.000
	AA2	H(1)=1.000	0.317
	AA3	H(1)=0.595	0.440
	MAN	H(1)=0.482	0.487
3	AA1	H(1)=0.057	0.811
	AA2	H(1)=0.000	1.000
	AA3	H(1)=0.238	0.626
	MAN	H(1)=1.485	0.223
4	AA1	H(1)=5.628	0.018
	AA2	H(1)=6.114	0.013
	AA3	H(1)=4.739	0.029
	MAN	H(1)=6.219	0.013
5	AA1	H(1)=0.000	1.000
	AA2	H(1)=0.056	0.814
	AA3	H(1)=0.134	0.715
	MAN	H(1)=1.534	0.215
6	AA1	H(1)=0.194	0.660
	AA2	H(1)=0.444	0.505
	AA3	H(1)=0.148	0.701
	MAN	H(1)=0.027	0.869
7	AA1	H(1)=0.115	0.734
	AA2	H(1)=1.208	0.272
	AA3	H(1)=0.307	0.579
	MAN	H(1)=0.051	0.821
8	AA1	H(1)=0.217	0.641
	AA2	H(1)=0.012	0.911
	AA3	H(1)=0.108	0.742
	MAN	H(1)=0.199	0.565
9	AA1	H(1)=0.319	0.572
	AA2	H(1)=0.701	0.403
	AA3	H(1)=0.054	0.816
	MAN	H(1)=0.618	0.432
10	AA1	H(1)=0.200	0.655
	AA2	H(1)=0.077	0.781
	AA3	H(1)=2.480	0.115
	MAN	H(1)=0.000	1.000
11	AA1	H(1)=0.217	0.641
	AA2	H(1)=0.196	0.658
	AA3	H(1)=0.148	0.700
	MAN	H(1)=0.048	0.827
12	AA1	H(1)=0.261	0.609
	AA2	H(1)=0.124	0.725
	AA3	H(1)=0.000	1.000
	MAN	H(1)=0.000	1.000

Table G-6: Kruskal-Wallis test results for the checklist for trust for the language condition

G-4 Frustration questionnaire

This section holds the answers participants gave to the frustration questionnaire presented at the end of the experiment. This questionnaire was only filled out by participants 6 to 21.

Participant 6

What frustrated you the most?

that I sometimes did not see conflicts ahead myself

Why did that frustrate you?

I like to be on top of things

Did you change your control strategy to reduce that frustration? In which way?

not really although I did look more at time slider I think

Participant 7

What frustrated you the most?

Agressiveness of warnings

Why did that frustrate you?

Took away freedom of coming op with own resolutions- e.g. unable to make own inputs and not allowing margins to become smaller

Did you change your control strategy to reduce that frustration? In which way?

It forced me to have the margins larger than achieved with my own "optimal" solutions.

Participant 8

What frustrated you the most?

The fact that there was only one solution option provided by the automation. Usually you'd steer towards a conflict solution long beforehand, but the automation decides that solution is not safe and wants to totally redirect a certain aircraft. That might be always playing it safe, but it also meant that the solution you'd been working towards for a few minutes just goes out the window. If you'd really panic and wouldn't know what to do, it is a sort of life line, but if a heading change is off by just a little bit, causing a conflict later on, it might overthrow your own intentions in some cases.

Why did that frustrate you?

Because you can sense that the system itself actually works and usually provides you with useful information, but it sometimes is disappointing that the automation deems your own solution, although not 100% perfect yet, to be irrelevant. I think this also lies in the fact that the automation only suggests one heading change in order to solve all issues, whereas your own solution might require two or three and then some correcting these as well. Therefore, from a safety perspective it's great if you don't know what to do anymore, but in terms of solving slightly more complicated issues or slightly adjusting the current situation in order to solve the conflict it might get in the way.

Did you change your control strategy to reduce that frustration? In which way?

Take a little less risk, i.e. applying slightly larger heading changes very far beforehand, such that there would be little chance of the automation trying to correct you. Also, if that were the case regardless, you'd have loads of time to apply a different heading change. It does require you to work a little faster but I really noticed the workload going down and the 'undesired' solution propositions occurring less frequently.

Participant 9

What frustrated you the most?

Not able to check which pair of aircraft corresponded to which circle

Why did that frustrate you?

Because you cannot take action in an early stage

Did you change your control strategy to reduce that frustration? In which way?

Yes, sometimes just change heading to check whether that particular aircraft corresponds to the circle

Participant 10

What frustrated you the most?

The intersections usually led me to make decision that were not "controller best practices".

Why did that frustrate you?

The situation is dependent on the number of a/c in the stream, any a/c behind/after the considered a/c and its speed. So it resulted in directing 1 a/c from each stream towards each other (though not conflicting).

Did you change your control strategy to reduce that frustration? In which way?

No. I felt this was still working better because I could make decisions based on future positions. This gave me a better awareness to deal with the selected maneuver.

Participant 13

What frustrated you the most?

when the automation took action for an aircraft at the sector border. it wanted to redirect it out of the sector and since I had to reject or accept the automation in an urgent situation I couldn't get a solution that keeps it inside the border without a conflict in time.

Why did that frustrate you?

because the time it took to think about rejecting the automation might have been enough to find a more suitable solution for the conflict.

Did you change your control strategy to reduce that frustration? In which way?

yes, i prevented to have to do too many changes close to the sector border.

Participant 14

What frustrated you the most?

Two scenarios: 1) My action restoring an aircraft to a course towards it target waypoint created a collision warning I did not expect, and immediately the automation did a suggestion that I could have easily solved myself. 2) The automation did a suggestion that was (probably) more along the best-practices, but was less future-proof than the plan that I had. If by accident a loss of separation warning would occur, the automation would counteract my plans.

Why did that frustrate you?

Already explained above

Did you change your control strategy to reduce that frustration? In which way?

In stead of estimating which new heading I would give an aircraft (with the risk of under-estimating and activating the automation), I mostly over-estimated and corrected in small steps. This led to more heading changes, but less collisions thus less automation intervention.

Participant 15

What frustrated you the most?

1.Sometimes clicking did not work. 2. Sometimes did not understand why the automation gave a resolution advisory. 3. Would like to see which dot belongs to which aircraft all the time, not just when the automation is active.

Why did that frustrate you?

1. Increases stress levels 2. Why would I accept it then? 3. Easier to predict which aircraft pairs will lead to a conflict

Did you change your control strategy to reduce that frustration? In which way?

Not really

Participant 16

What frustrated you the most?

Not being able to click on the rounds in the separation monitor display when the automation was turned off, to see which aircraft were in conflict.

Why did that frustrate you?

I could make an estimate of which aircraft were possibly in conflict, but I was not always sure and could not check my ideas with that display.

Did you change your control strategy to reduce that frustration? In which way?

I looked more ahead from the start and started to look at other combinations far ahead when the automation was turned on.

Participant 17

What frustrated you the most?

1. That I received a warning but was not immediately clear what the problem is and how automation wants to solve that problem. 2. Interface inconsistency: 'automation on' is on the left side while the 'accept-reject' bar appears on the right side of the screen

Why did that frustrate you?

1. When receiving a warning, I wanted to resolve the problem as fast as possible, but getting confused by the system's proposal delayed my reaction. 2. I had to check both corners to ascertain what the state of the system was.

Did you change your control strategy to reduce that frustration? In which way?

1. I tried to plan ahead and avoid receiving too many proposals from automation. Also, when I did receive a warning / proposal, I was more prone to implement my own rapid solution instead of taking 1-2 more seconds to figure out the intention of automation. Or I just immediately accepted the proposal, but reassessed the solution afterwards. 2. I just tried to be aware of both corners.

Participant 19

What frustrated you the most?

That when the automation poped out I could not ignore it

Why did that frustrate you?

Because the alert made me aware of my error and I wanted to be able to solve it myself

Did you change your control strategy to reduce that frustration? In which way?

Yes, I just accepted everything the automation suggested

Participant 20

What frustrated you the most?

Sometimes an overflow of conflicts occurred (2 or 3 at the same time), which makes it tempting to use the automated solver to solve them all.

Why did that frustrate you?

Because it solves the problem, but sometimes with a different strategy than you had in mind. So you have to implement a different strategy than intended.

Did you change your control strategy to reduce that frustration? In which way?

Yes, I started to realize how the system would solve the problem and I adapted my own strategy to that.

Participant 21

What frustrated you the most?

The fact that for some long term conflicts I wanted to resolve with a different RA than what the adaptive automation proposed I had to reject the proposed RA and first decide if the proposed RA made sense, hence losing time. But this is in the case of long term strategy that I thought would work and not necesarrily would work

Why did that frustrate you?

Because I sort of lost the sense of control/decision that would allow me to reason further for my strategy to maintain separation.

Did you change your control strategy to reduce that frustration? In which way?

For some complicated conflicts I decided to let the automation do whatever it wants and see from there, because I was already spending too much time thinking about the next step when obviously some other further conflicts where bound to happen anyway. I also changed the cross conflict strategy by also vectoring the faster aircraft and the slower one, and not only the slower one. However that did not seem always to prove useful.

Bibliography

Al-Jabri, I., & Roztocki, N. (2015). Adoption of erp systems: Does information transparency matter? *Telematics and Informatics*, *32*, 300-310.

Anderson, J. (1988). The expert module. In M. Polson & J. Richardson (Eds.), *Foundations of intelligent tutoring systems* (p. 21-54). Lawrence Erlbaum Associates.

Anderson, J., Boyle, C., & Reiser, B. (1985). Intelligent tutoring systems. *Science*, *228*, 456-462.

Bessiere, K., Ceaparu, I., Lazar, J., Robinson, J., & Schneiderman, B. (2004). Social and psychological influences on computer user frustration. In E. Bucy & J. Newhagen (Eds.), *Media access: Social and psychological dimensions of new technology use* (p. 91-103). Lawrence Erlbaum Associates.

Billings, C. (1997). *Aviation automation: The search for a human-centered approach* (1st ed.). 10 Industrial Avenue, Mahwah, NJ 07430-2262 USA: Lawrence Erlbaum Associates.

Bolic, T. (2006). *Automation adoption and adaptation in air traffic control*. Unpublished doctoral dissertation, UNIVERSITY OF CALIFORNIA, BERKELEY.

Bubb-Lewis, C., & Scerbo, M. (1997). Getting to know you: Human-computer communication in adaptive automation. In M. Mouloua & J. Koonce (Eds.), *Human-automation interaction: Research and practice* (p. 92-99). Lawrence Erlbaum Associates.

Chen, J., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awarenessbased agent transparency* (Technical Report No. ARL-TR-6905). Aberdeen Proving Ground, MD 21005-5425: U.S. Army Research Laboratory.

Christoffersen, K., & Woods, D. (2002). How to make automated systems team players. In *Advances in human performance and cognitive engineering research, volume 2* (p. 1-12). Elsevier Science Ltd.

Corbett, A., Koedinger, K., & Anderson, J. (1997). Intelligent tutoring systems. In M. Helander, T. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction, second, completely revised edition* (p. 849-874). Elsevier Science B.V.

Cramer, H., Evers, V., Ramlal, S., Someren, M. van, Ruthledge, L., Stash, N., et al. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User modeling and user-adapted interaction*, *18*, 455-496.

DSNA, & ENAV. (n.d.). *Erato electronic environment.* `http://dsnaservices.com/wp-content/uploads/2015/03/ERATO.pdf`.

Edwards, M., Fuller, D., & Vortac, O. (1995). The role of flight progress strips in en route air traffic control: a time-series analysis. *Human-Computer Studies*, *43*, 1-13.

ENAV. (n.d.). *Erato.* `https://www.enav.it/enavWebPortalStatic/ATM2016/internationalprogrammespartnerships.html`.

Endsley, M. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (p. 163-181). Lawrence Erlbaum Associates.

Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *ERGONOMICS*, *42*(3), 462-492.

Endsley, M., Sollenberger, R., & Stein, E. (1999). The use of predictive displays for aiding controller situation awareness. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 43, p. 51-55).

Eurocontrol. (2010). *Eurocontrol specification for medium-term conflict dectection.*

Ghazizadeh, M., Lee, J., & Boyle, L. (2012). Extending the technology acceptance model to assess automation. *Cogn Tech Work*, *14*, 39-49.

Göritzlehner, R., Borst, C., Ellerbroek, J., Westin, C., van Paassen, M., & Mulder, M. (2014). Effects of transparency on the acceptance of automated resolution advisories. *IEEE International Conference on Systems, Man, and Cybernetics*, 2965-2970.

Grunwald, A., Robertson, J., & Hatfield, J. (1980). *Evaluation of a computer-generated perspective tunnel display for flight-path following* (Technical Report No. L-13253). NASA Langley Research Center.

Hilburn, B., Molloy, R., Wong, D., & Parasuraman, R. (1993). Operator versus computer control of adaptive automation. *Seventh International Symposium on Aviation Psychology*, 31-36.

Höök, K. (2000). Steps to take before intelligent user interfaces become real. *Interacting with computers*, *12*(4), 409-426.

IJtsma, M. (2016). *Adaptive automation based on air traffic controllers decision-making.* Unpublished M.Sc. Thesis, Faculty of Aerospace Engineering, Delft University of Technology.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of cognitive task design* (p. 147-169). Lawrence Erlbaum Associates.

Inagaki, T. (2008). Smart collaboration between humans and machines based on mutual understanding. *Annual Reviews in Control*, *32*, 253-261.

International Civil Aviation Organization. (1996). *Rules of the Air and Air Traffic Services* (Tech. Rep. No. 4444-RAC/501). Montreal.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53-71.

Kaber, D., & Endsley, M. (2004). 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), 113-153.

Kaber, D., Riley, J., Tan, K., & Endsley, M. (2001). On the design of adaptive automation for complex systems. *INTERNATIONAL JOURNAL OF COGNITIVE ERGONOMICS*, *5*(1), 37-57.

Klomp, R., Mulder, M., Van Paassen, M., & Roerdink, M. (2011). Re-design of an inbound planning interface for air traffic control.

Lee, J., & Moray, N. (1994). Trust, self-confidence, and operator's adaptation to automation. *Int. J. Human-Computer Studies*, *40*, 153-184.

Lee, J., & See, K. (2004). Trust in automation: Designing for appropriate reliance. *HUMAN FACTORS*, *46*(1), 50-80.

Lewandowski, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, *6*(2), 104-123.

Lin, C., Yenn, T., & Yang, C. (2010). Evaluation of operators performance for automation design in the fully digital control room of nuclear power plants. *Human Factors and Ergonomics in Manufacturing and Service Industries*, *20*(1), 10-23.

Maslow, A. (1941). Deprivation, threat and frustration. *Psychological Review*, *48*, 364-366.

McCarthy, J., Miles, V., & Monk, A. (1991). An experimental study of common ground in text-based communication. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 209-215.

Mercado, G. (2013). *Context-aware adaptive automation in atc.* (Literature review)

Miller, C. (2014). Delegation and transparency: Coordinating interactions so information exchange is no surprise. In R. Schumaker & S. Lackey (Eds.), *Virtual, augmented and mixed reality. designing and developing virtual and augmented environments* (p. 191-202). Springer International Publishing.

Mosier, K., & Skitka, L. (1996). Human decision makers and automated decision aids: Made for each other? In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (p. 201-220). Lawrence Erlbaum Associates.

Mouloua, M., Smither, J., Vincenzi, D., & Smith, L. (2015). 9. automation and aging: Issues and considerations. *Advances in Human Performance and Cognitive Engineering Research*(2), 213-237.

Muir, B. (1987). Trust between humans and machines, and the design of decision aids. *Int. J. Man-Machine Studies*, *27*, 527-539.

Norman, D. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'overautomation'. *Philosophical transactions of the royal society B*, *327*(1241), 585-593.

Oduor, K., & Wiebe, E. (2008). The effects of automated decision algorithm modality and transparency on reported trust and task performance. *PROCEEDINGS of the HUMAN FACTORS AND ERGONOMICS SOCIETY 52nd ANNUAL MEETING*.

Onnasch, L., Wickens, C., Li, H., & Manzey, D. (2013). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *HUMAN FACTORS*, *20*(10), 1-13.

Parasuraman, R. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230-253.

Parasuraman, R., Bahri, T., Deaton, J., Morrison, J., & Barnes, M. (1990). *Theory and design of adaptive automation in aviation systems* (Technical Report No. N62269-90-0022-5931). Washington, DC: Cognitive Science Laboratory, THE CATHOLIC UNIVERSITY OF AMERICA.

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICSPART A: SYSTEMS AND HUMANS*, *30*(3), 286-297.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. *Computers and Mathematics with Applications*, *33*(5), 128.

Riley, V. (1996). What avionics engineers should know about pilots and automation. *IEEE Aerospace and Electronic Systems Magazine*, *11*(5), 3-8.

Roberts, A., & Pember, S. (2007). *Air traffic control* (Nos. US 8,255,147 B2).

Scerbo, M. (1996). Theoretical perspectives on adaptive automation. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance* (p. 37-63). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

SESAR. (n.d.). *Sector team operations - en-route air traffic organiser.* `http://www.sesarju.eu/sesar-solutions/conflict-management-and-automation/sector-team-operations-en-route-air-traffic`.

Sheridan, T., & Parasuraman, R. (2005). Human-automation interaction. *Reviews of human factors and ergonomics*, *1*(1), 89-129.

Sheridan, T., & Verplank, W. (1978). *Human and computer control of undersea teleoperators* (Technical Report No. N00014-77-C-0256). Cambridge, MA 02139: Massachusetts Institute of Technology.

Thurman, D., Brann, D., & Mitchell, C. (1997). An architecture to support incremental automation of complex systems. In *Proceedings of the 1997 ieee international conference on systems, man and cybernetics* (p. 1174-1179).

Tomizuka, M., & Whitney, D. (1976). The human operator in manual preview tracking (an experiment and its modeling via optimal control). *Journal of Dynamic Systems, Measurement, and Control*, *98*(4), 407-413.

Westin, C., Borst, C., & Hilburn, B. (2015). Strategic conformance: Overcoming acceptance issues of decision aiding automation? *IEEE Transactions on Human-Machine Systems*, *46*(1), 41-52.

Wickens, C. (1998). Automation in air traffic control: The human performance issues. *Third International Conference on Automation Technology and Human Performance*.

Wickens, C., & Morphew, E. (1997). *Predictive features of a cockpit traffic display: A workload assessment* (Technical Report No. ARL-97-6/NASA-97-3). NASA Ames Research Center.

Woods, D. (1984). Visual momentum: a concept to improve the cognitive coupling of person and computer. *International Journal of Man-Machine Studies*, *21*(3), 229-244.

Woods, D., & Patterson, E. (2000). How unexpected events produce an escalation of cognitive and coordinative demands. In P. Hancock & P. Desmond (Eds.), *Stress workload and fatigue.* Lawrence Erlbaum Associates.

Woods, D., & Sarter, N. (1998). *Learning from automation surprises and "going sour" accidents: Progress on human-centered automation* (Technical Report No. NCC 2-592). NASA Ames Research Center.