AIR TRAFFIC CONTROLLER DECISION-MAKING CONSISTENCY AND CONSENSUS IN CONFLICT SOLUTION PERFORMANCE

Carl Westin
Delft University of Technology
PhD student
Kluyverweg 1, 2629 HS Delft, The Netherlands
C.A.L.Westin@tudelft.nl
Brian Hilburn (CHPR), Clark Borst (Delft University of Technology)

ABSTRACT

Consistency and consensus in conflict solution performance was investigated in two real-time simulations involving experienced air traffic controllers and trainees. The simulations consisted of participants repeatedly playing a specific en route traffic scenario. Conflict solution performance and consistency was measured by analysing participant's solutions to a specifically designed conflict. Consensus was determined by comparing participant's consistent conflict solution behaviour. Four different clusters of consistent behaviour were identified. All participants could be classified as consistent according to one or more of these clusters. Consensus in solving the designed conflict could not be determined between participants. There were, however, smaller groups of participants who solved the designed conflict similarly. Participants varied considerably in terms if when to interact and solve the conflict. This is a potential issue when it comes to determining thresholds for the timing of decision-aid interaction.

1 INTRODUCTION

With more complex and demanding task-sharing and decision-making tasks, the automation's ability to consider human variability should increase. Automation will have to better acknowledge individual differences, not only in terms of human performance variables such as workload, stress, and fatigue, but also in decision-making and problem solving [1]. Technology designers have, for quite some time, considered individual physical differences, based on anthropometric data. An example can be found in the individual seat memory feature of modern cars. Another category of individual characteristics can be tied to our perception and sensory systems. For example, various settings of computers and smartphones can be adjusted for individual preferences. Increasing font sizes and volume can compensate age-related deteriorations in vision and hearing. Background pictures and ring tones are examples of individual preferences that can be considered more aesthetic in nature, independent from anthropomorphic configuration and perceptual preferences. A fourth category of individual design characteristics can be tied to differences is our cognitive style, or problem solving and decision-making preferences [2]. One example is the increasing use of recommendation systems for content in museums, art exhibitions, and other cultural heritage domains. Here adaptive automation (which dynamically modifies its assistance based on the inferred real-time needs of the user) is used to generate recommendations on cultural content that better fit the individual visitor [3].

This type of individually sensitive automation is especially interesting for conflict detection and resolution (CD&R) automation in air traffic control (ATC). In current ATC, however, individualised operator-adaptive automation is non-existing. Although controllers can, to some extent, personalise the radar environment interface, this customisation is usually limited to turning information off or on. Generally, ATC research and automation design have assumed air traffic controllers to be *homogeneous* in their CD&R performance. As such, CD&R automation has traditionally been designed to fit a single stereotype air traffic controller (i.e. one size fits all).

ATC research has predominantly focused on investigating the degree of consensus (inter-rater reliability) between controllers, while giving little attention to consistency (intra-rater reliability). While researchers generally agree that controllers are homogeneous [4-5] it is also widely acknowledged that performance and work practices differ significantly between countries and even control centres in the same given country [6]. More recently, the MUFASA project showed that controllers differed in CD&R performance, and that they could be assigned different conflict solution styles [7]. Based on controllers own solutions to conflicts (as recorded in a manual prequel simulation), a decision-aid was developed that would suggest resolution advisories to detected conflicts. In a subsequent large-scale real-time simulation, the decision-aid supported controllers in CD&R. Unbeknownst to the controllers, 50% of all resolution advisories received were replays of their own previously recorded conflict solutions (called conformal advisories). The other half consisted of (adequate but qualitatively different) solutions made by other participants (called non-conformal advisories). While controllers accepted and agreed more with conformal advisories than non-conformal advisories, they also rejected and disagreed with their own solutions in almost 25% of the cases. Possibly, this finding can be attributed to internal variability, and that controllers are simply inconsistent in their conflict solution decision-making strategies over time.

Research evidence on intra-controller consistency has not provided a coherent picture. There is evidence that controllers are consistent [8] and inconsistent [9] in traffic complexity judgements. In a study using the Cochran-Weiss-Shanteau (CWS) index to measure discrimination and consistency [10], controller performance inconsistencies increased with complexity [11]. There seems to have been insufficient research into the topic to draw clear conclusions, and further work is needed to determine the causes and extent of such inconsistency, and the potential implications for automation design.

The two main research questions were:
- Consistency: Are controllers internally consistent in their conflict resolution performance over time?
- Consensus: Do controllers solve conflicts similarly?

The aim of this study was to qualitatively investigate controllers' degree of consistency and consensus in conflict solution performance. Section 2 describes the method and experimental design. Section 3 presents the results of real-time simulations. Section 4 discusses the relevance of consistency and consensus in CD&R performance.

2 METHOD

Two human-in-the-loop simulations were conducted in which participants' performance in conflict avoidance was recorded and analysed. Participants played multiple manual runs of the same, repeated (unrecognisable) traffic scenario. Specifically, their performance was analysed to determine consistency and consensus in relation to a specifically designed conflict. The experimental design and simulator, including traffic scenarios, was built on previous work conducted within the MUFASA project [7].

2.1 Participants

A total of fourteen volunteers with ATC experience participated in the simulation. Two simulations, at separate locations, were conducted with different groups of participants. Nine participants (three females and six males) at Malmö Air Traffic Control Center were Swedish trainees who had finished their basic training and about to start the final on-the-job-training. Trainees were between 24 and 29 years old (mean = 26 years). Although basic training had been similar for all trainees, eight were specifically trained for Terminal Manoeuvring Area control, while one was trained for a tower position. Five participants were experienced controllers from Norrköping Terminal Control Center, Sweden. Age varied between 26 and 47 (mean = 32,8 years). Experience varied between thirteen months and 24 years (mean = 8,7 years).

2.2 Apparatus

The simulation ran on a portable computer connected to an external 21" screen with a resolution of 1600x1200 pixels. The simulator was a Java-based application using OpenGL extensions. Participants interacted with the simulator through mouse and keyboard. A post simulation questionnaire was collected through an online survey using Google Forms. The simulator ran at 2x real speed. Aircraft plots on the display were updated every second to simulate a 1 Hz radar update frequency. The simulation program recorded and logged all scenario activities automatically. This allowed for export of data to desired software and subsequent data analysis.

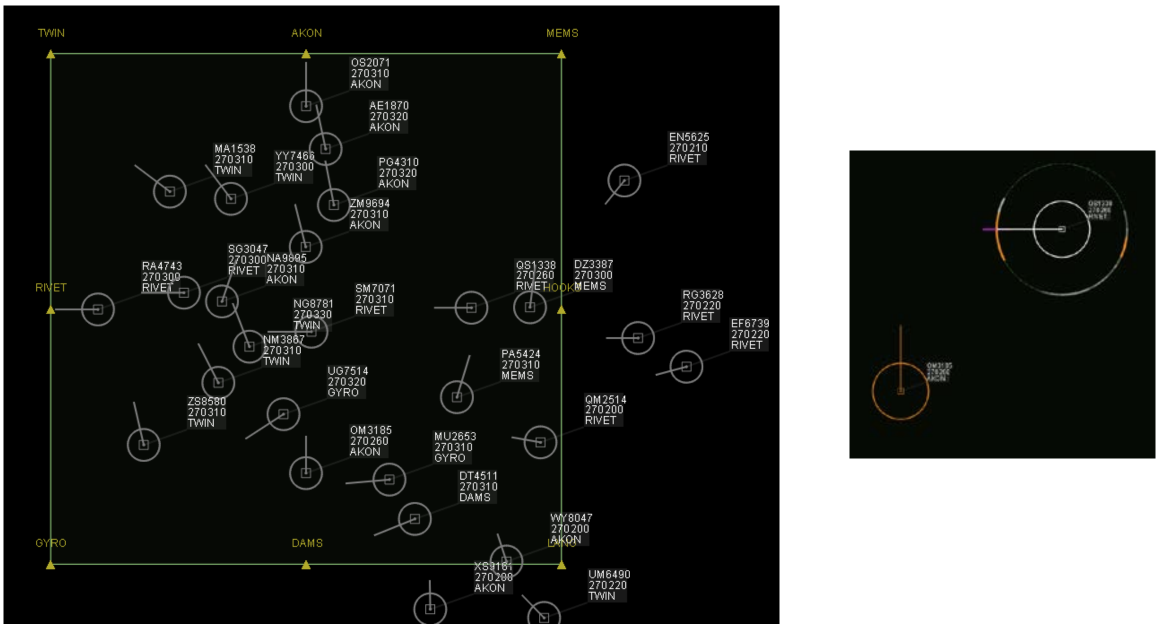
2.3 Traffic scenarios

The simulation comprised a series of short (two minute) traffic scenarios of level en-route traffic. All scenarios consisted of a hypothetical squared en route sector, 80 x 80 NM in size (Figure 1a). One measurement scenario and three dummy scenarios were created. The measurement scenario was repeated four times, while each dummy scenario was repeated twice. The purpose of dummy scenarios was to conceal the repetitions of the measurement scenario, and prevent scenario recognition. If a participant recognises a scenario and designed conflict, the participant may try to recall a previous response (solution) rather than coming up with an independent solution. In addition, scenarios were mirrored to further mitigate the risk of recognition. Although scenario mirroring or rotations provide a risk for introducing confounds, it is a commonly used technique in ATC simulations to prevent scenario recognition in repeated measures designs (e.g. [12]).

The measurement scenario contained a specific designed conflict against which participants' conflict solutions could be analysed. The designed conflict was created to occur at right angles (90 degrees) in the exact middle of the sector between aircraft A and aircraft B (Figure 1b). The conflict pair was initially aligned to the exit points and required no initial interaction. Both aircraft were equally far away from the closest point of approach (set to 0 nm), travelling at the same speed of 260 kts indicated airspeed in zero wind conditions. Additionally, both aircraft had identical speed envelopes (ranging from 200 to 320 kts). Unless participants' intervened, the aircraft would lose separation (5 nm) 104 seconds after the scenario started. The conflict geometry and aircraft configurations were selected to mitigate biased solutions. By simply looking at the relationship between the conflict pair, both aircraft had identical solution possibilities.

"Noise" aircraft were used to increase scenario complexity, complicate and prevent early conflict detection (of the designed conflict), and make scenarios more realistic. Noise aircraft were configured so that, even if participants interacted with them, their presence would not interfere with the conflict pair. Sufficient margins were provided to allow the designed conflict to be solved with both large left or right vectors in combination with any desired speed changes. Aircraft ahead of the conflict pair were given higher cruising speeds, while aircraft behind were given lower cruising speeds.

The ATC environment was presented as a futuristic sector and workstation. As such, current decision aiding tools that might have been familiar to the participants were not included. The interface was based on a modified prototype of the Solution Space Diagram (SSD) under development at Delft University of Technology [13]. The SSD is the result of an ecological interface design approach to an air traffic controller's work domain and the task of CD&R. For our simulations, the SSD was used as a tactical decision-aiding tool for separation assurance and conflict solving. The SSD facilitated the constraints and solution possibilities of the selected aircraft in relation to intruder aircraft by providing colour coded GO and NO-GO areas. In Figure 1b the SSD of aircraft A indicates a conflict (amber area in SSD ahead of QS1338) with aircraft B (in amber). A vector to the right or left by roughly 30 degrees or more would have prevented a separation loss. The SSD appeared automatically when a participant selected an aircraft.



a *b*

Figure 1: (a) Measurement scenario. (b) Designed conflict between aircraft A (QS1338) and aircraft B (OM3185).

Sector throughput was considerably higher than current day. Furthermore, an open airspace with only sector entry and exit points was used to imply a free-routing environment. Finally, participants were instructed that whenever an aircraft was interacted with, the clearance was automatically communicated to the aircraft through data link. For purposes of experimental control, traffic was restricted in level flight operating at FL 270. All traffic consisted of short- to medium-range narrow-body twinjet aircraft (such as the Boeing 737 and Airbus 320 families). Meteorological conditions were held constant throughout scenarios (e.g. no wind).

2.4 Experimental design

The consistency study is a type of single-subject design specifically investigating the variability in participant's behavioural responses to a specific stimulus over time. We used a repeated measures design with measurement scenario, and specifically the designed conflict, as a single independent variable. Each scenario was repeated four times per participant. Dependent variables comprised all aspects defining the conflict solution implemented. As such, consistency was analysed qualitatively by investigating participants' performance and actions and identifying similarities in the four separate solutions of the designed conflict.

2.5 Procedures

Participation lasted roughly one and a half hours and included a simulation part and a questionnaire part. Following briefing and consent procedures, we conducted 14 training runs before starting the measurement session consisting of 10 scenarios. While the order and timing of the four repeated measurement scenarios were fixed in all playlists, the order of dummy scenarios varied between participants according to a Latin square design.

There were two main tasks: to resolve conflicts, and clear aircraft to their exit points as indicated in their flight label. To keep participants motivated and focused, feedback on task efficiency was provided through a continuously updated performance score. Participants were supported by a short-term conflict detection system that would provide a warning whenever a conflict was detected. Up to 60 seconds prior to a separation loss, the involved aircraft would turn amber. With less than 30 seconds before separation loss, an auditory alert was provided and the involved aircraft were displayed in red. After each scenario, participants were asked to indicate their perceived scenario difficulty on a 1-100 scale. Before continuing with the next scenario, performance feedback was provided in terms of an average score. After the simulation, participants filled out a short questionnaire.

3 RESULTS

Consistency was measured by determining the similarity across the four scenario repetitions in participants' (a) overall scenario performance and (b) solutions to the designed conflict. Conflict solution behaviour was categorised into different control action styles. Identical behaviour across all four repetitions represented high consistency. In contrast, large variations in control actions indicated an inconsistent behaviour. Consensus was analysed by comparing participant's control action styles for (a) scenario performance data and (b) solutions to the designed conflict. Furthermore, the two different samples of trainees and experienced controllers allowed for consistency analysis between groups.

As a first step, participant's control actions were assessed against a solution parameters hierarchy whereby the solution was examined through its parts (i.e. time of intervention, aircraft choice, resolution type, and vector direction). Control actions were then assessed by graphically inspecting recordings of prequel data, specifically looking at the conflict solution geometry (i.e. relationship between the conflict pair).

3.1 Overall scenario performance

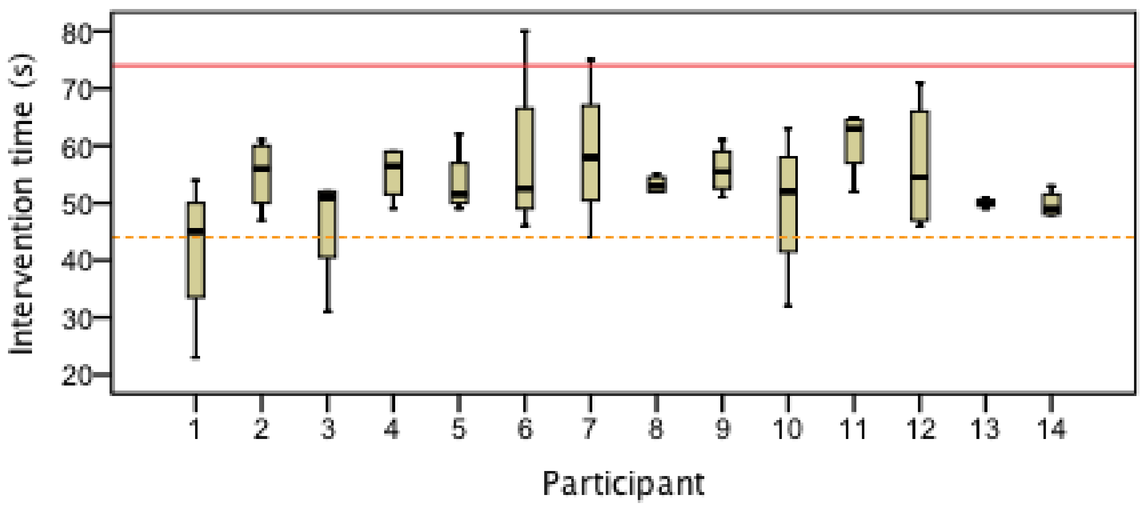
This data refers to the four repetitions of the measurement scenario and does not include dummy scenario performance. Number of SSD inspections was greater among trainees (*mean* = 39.6, *SD* = 8.7) than experienced controllers (*mean* = 33.1, *SD* = 4.3). Experienced controllers made more interactions per scenario (*mean* = 19.5, *SD* = 3.1) than did trainees (*mean* = 16.8, *SD* = 2.6). The most frequently used control action was heading commands (trainees: *mean* = 14.9, *SD* = 2,6 experienced controllers: *mean* = 16.7, *SD* = 3.2). Speed was used slightly more frequently than combinations of speed and heading, although neither was common. Analysis of the scenario performance suggests that participants overall were homogenous in terms of used control actions. The perceived scenario difficulty varied among participants with experienced controllers perceiving scenarios slightly less difficult (*mean* = 45.8, *SD* =

14.3) than trainees (mean = 54.4, *SD* = 13.4). Both experienced controllers and trainees found scenarios neither very low nor very high in complexity.

3.2 Intervention time

Intervention time was measured from the start of the scenario until the first control action undertaken to solve the designed conflict. Intervention time was more variable between trainees than between controllers. For trainees, intervention time ranged from 23 to 80 sec from scenario start (*mean* = 53.0s, *SD* = 9.8s), and for controllers, intervention time ranged from 32 to 71 sec (*mean* = 53.4s, *SD* = 8.7s). In relation to these numbers it should be noted that the caution alert (designed conflict aircraft change in colour to amber to indicate conflict) was provided 44 seconds after the scenario started. The warning alert (both aircraft turned red and a continuous auditory alert sound was provided) was triggered 74 seconds after the scenario started. The majority of interventions occurred between the caution and warning alert for both trainees (88.8%) and controllers (95%).

Across the four scenario repetitions, consistency in intervention time clearly differed between participants. As shown in Figure 2, two participants were very consistent with interventions taking place at about the same time in all four repetitions (within a four and three second interval, respectively). Eight out of fourteen participants intervened within the same 15-second window in all four repetitions. The other six participants intervened within a larger timeframe of 22 to 35-seconds.



X-axis reference lines: amber (dashed) = warning alert 44s; red (solid) = caution alert

Figure 2: Boxplot showing intervention times for all participants across scenario repetitions.

3.3 Solution parameter hierarchy analysis

A decision tree analysis of the control actions used to solve the designed conflict was conducted. Note that only first interactions were considered in this analysis. Trainees had a slight preference for interacting with aircraft A (21 counts) over aircraft B (15 counts), whereas controllers split evenly between the two aircraft (10 counts each). For both groups, the most common aircraft choice distribution (across the four scenario repetitions) was to select aircraft A twice and aircraft B twice (8 out of 14 participants).

Heading commands were the most frequently used interaction to solve the designed conflict. Out of 36 scenarios, trainees solved 31 of the designed conflicts with heading commands. Only five solutions consisted of heading and speed combinations, all for aircraft A. A similar pattern is found for experienced controllers, who solved the conflict 19 out of 20 times by implementing a vector. Only once was a combination implemented.

Looking at the resolution type direction data, participants in both groups differed in their preference for the direction of the vector given. While trainees preferred right vectors regardless of aircraft (67% for aircraft A and 73% for aircraft B), experienced controllers were split between vectoring aircraft A left (70%) and aircraft B right (80%). As such, controllers generally vectored the controlled aircraft behind the intruder aircraft. Conflict solutions based on combinations of speed and heading were rare (11% in total and all made with aircraft A). Individually, trainees displayed a stronger tendency for using right vectors. Five trainees used only right vectors, or a combination of three right vectors and one left vector, to solve the four repetitions of the designed conflict. In contrast, experienced controllers were more evenly spread between right and left vectors and there was no tendency for a directional preference.

3.4 Graphical analysis of conflict solutions

Qualitative analysis of the SSD spider-graph diagrams illustrated different consistent patterns among participants. Three participants (all trainees) consistently chose the same aircraft in all four repetitions. In Figure 3, SSD spider-graphs from two of these participants with contrasting consistent behaviour are shown. Whereas one participant preferred to interact with aircraft A, the other chose aircraft B. Note that both participants preferred a right vector solution.

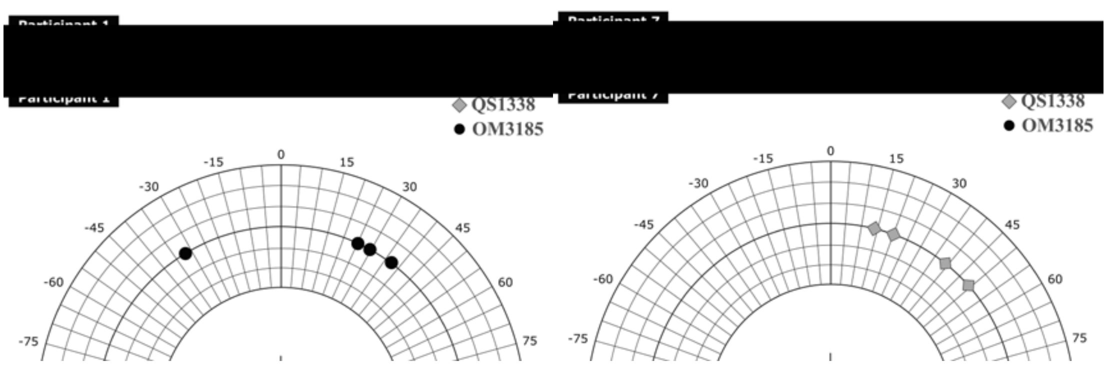


Figure 3: SSD spider-graph diagrams showing consistency in choice of aircraft.

The squared spider-graph diagram in Figure 4 illustrates another potential consistency pattern. This participant solved the conflict by always having aircraft B go behind A. The consistent behaviour was, however, accomplished by twice interacting with aircraft B and twice with aircraft A. In terms of solution geometry, the consistency analysis looked at the resulting traffic pattern of the conflict solution. Specifically, it identified which of the two conflict aircraft went behind the other. Out of the eight participants who were found to be consistent according to this classification, all except one consistently made aircraft B go behind A. Note that this is the same as aircraft A going in front of B, and that this consistency classification does not discriminate between which aircraft the interaction was made with.

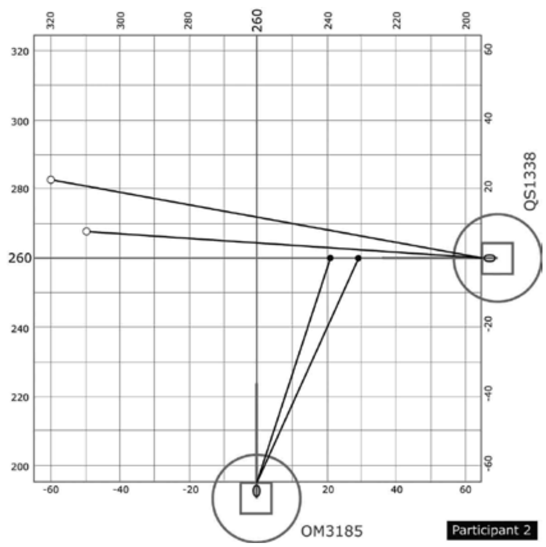
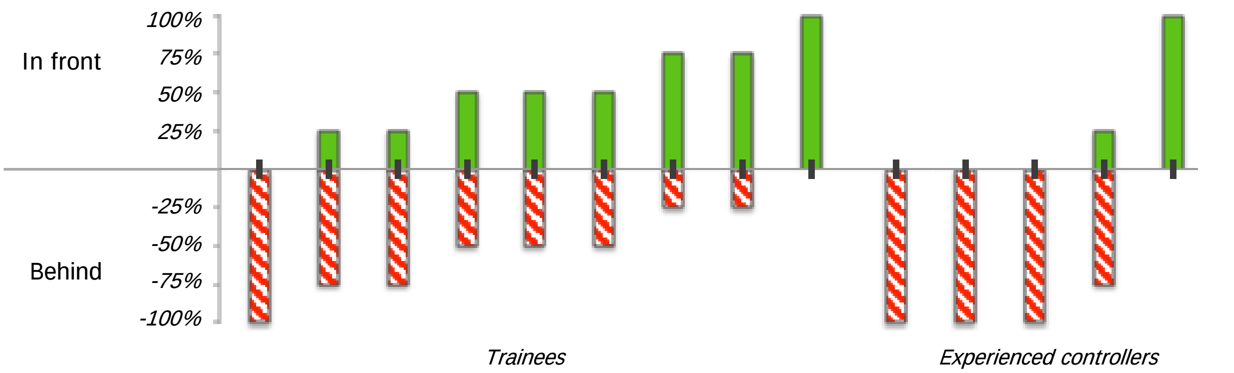


Figure 4: Squared spider-graph diagram showing solution pattern consistency.

3.5 Control problem analysis

Conflict solutions were analysed from a simple dichotomous control problem perspective, independently of the solution parameter hierarchy. In this view, the designed conflict constitutes a control problem whereby an intruder aircraft poses a threat to the safe flightpath of the controlled aircraft. The airspace ahead of the controlled aircaft is constrained by boundaries for separation assurance generated by the intruder aircraft. The control problem can only be solved by turning the controlled aircraft either behind or in front of the intruder aircraft, and analysis relied on classifying the direction of initial turn (note that multiple interventions were possible). This analysis showed that the conflict solutions of eleven participants could be consistently described in this way. Four participants consistently vectored the controlled aircraft in front of the intruder, while seven did the opposite.

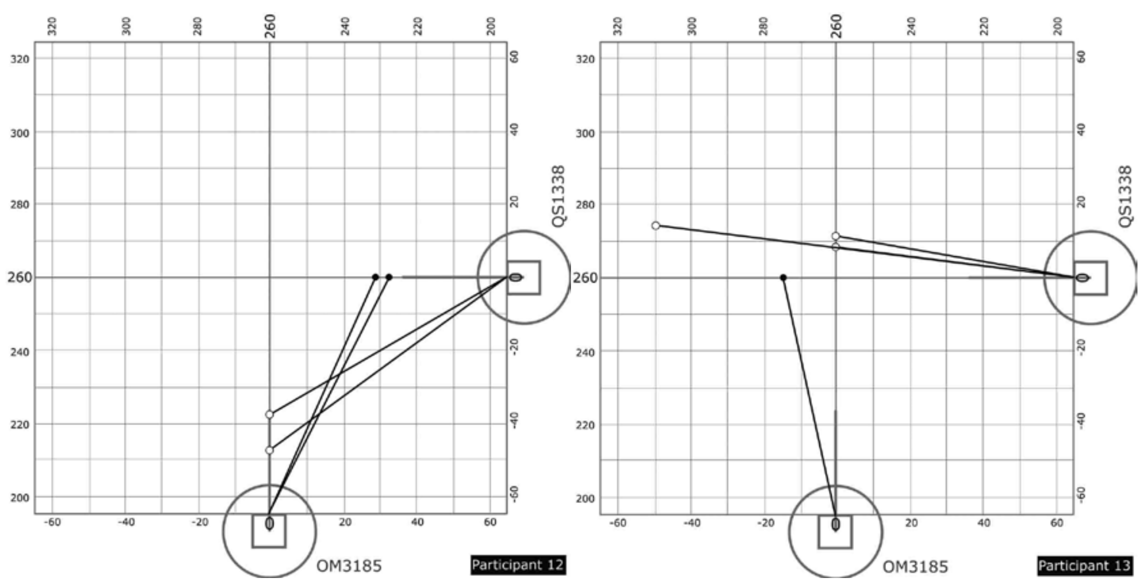
Bar diagrams of the control problem analysis is shown in Figure 5. Each bar represents the distribution of the control problem analysis per participant across the four scenario repetitions. E.g. participants with equally long bars (50% green and 50% red/white barber poles) on either side of the X-axis reference line solved the designed conflict by twice vectoring the controlled aircraft behind, and twice in front the intruder aircraft. Bars reaching 100% indicate a highly consistent behavior, since repeated this solution to the control problem in all four repetitions. Six participants were highly consistent (all four scenarios solved in the same way), although at opposite sides of the scales. Note that all except one experienced controller was found consistent according to this classification. As seen in Figure 5, two participants very consistently vectored the controlled aircraft in front of the intruder aircraft. In contrast, four participants consistently vectored the controlled aircraft behind the intruder aircraft. A slightly less consistent behaviour was observed for five participants. Three of these participants chose to vector the controlled aircraft behind the intruder aircraft in three repetitions, whereas two participants did the opposite. Three participants (all trainees) displayed the least consistent behavior in that they vectored the controlled aircraft in front equally many times as they vectored it behind.



X-axis: Participants per group and in order of consistent behaviour.
Y-axis: Degree of consistency.

Figure 5: Bar diagram of control problem analysis

For experienced controllers, the control problem analysis revealed a more consistent solution behavior compared to the trainee data (Figure 5). To further investigate this, squared spider-graphs were created. These graphs better illustrate participant's conflict solutions and the resulting relationship between the conflict aircraft. Out of the four participants who consistently turned the controlled aircraft behind the other intruder aircraft, three did so by twice vectoring aircraft A behind B, and twice vectoring aircraft B behind A (Figure 6a). One participant vectored aircraft B behind A in three scenarios, and aircraft A behind B in one scenario. The two participants who consistently turned the controlled aircraft in front of the intruder aircraft differed slightly in terms of which aircraft was chosen. One participant chose to always vector aircraft A in front of B, while the other did so in three out of four scenarios (Figure 6b)



a *b*

Figure 6: Squared spider-graph diagrams showing control problem consistency.

3.6 Conflict solutions involving multiple interactions

Participants often solved the designed conflict by interacting with the conflict pair multiple times. In total, this occurred in almost 40% of all scenarios played (22 out of 56 for all participants). Among experienced controllers, six out of twenty scenarios (30%) were solved with two (15%) or three (15%) control actions. This solution preference was even stronger among trainees, with sixteen out of thirty-six scenarios (almost 45%) of which 12 (33%) involved two interactions, and 4 (11%) three interactions. A closer inspection of these occurrences revealed that three participants consistently used multiple interactions to solve the designed conflict. For example, one participant first interacted with aircraft A, vectoring it right (in front of B). Second, the participant interacted with aircraft B, vectoring it right in two occasions, decreasing the speed once, and vectoring it left once.

We identified three reasons underlying these multiple interactions. First, some participants preferred to vector both aircraft a little, rather than implementing a large deviation to only one aircraft. In these cases, the dual interaction can be interpreted as a deliberate solution strategy. Second, a few instances were likely the result of a failed conflict solution for which the participant tried to salvage the situation by re-solving the conflict (in either one interaction, or another two interactions). Finally, participants might have missed or misunderstood the light grey NO-GO area in the SSD interface, believing that a tighter vector was possible. In the SSD, a light grey area indicated a long-term conflict warning (1 minute to infinite ahead in time). After vectoring an aircraft into a light grey area, the participant incorrectly believed the conflict to be solved. A closer investigation of the intervention time for the second, and sometimes third, interaction provided some insight to the latter two potential explanations. A short time period between interactions suggests a deliberate strategy. In thirteen out of totally fifty-six scenarios dual interactions were made within a ten second time-window, suggesting that both interactions were part of a deliberate strategy. During this time window, participants did not interact with any aircraft other than the conflict pair, suggesting that they were focused on solving the designed conflict. In several situations, the first aircraft interacted with had not reached its new state before second interaction was initiated.

3.7 Groups of consistent conflict solution behaviour

In summary, the analysis of participant's conflict solutions (of the designed conflict) generated four different groups of consistency. Each participant was, at least once, assigned a consistent behaviour according to a consistency group. A consistent group was defined as a certain control action pattern being implemented in three or more (out of four) scenario repetitions. A consistent conflict solution, as defined by the solution parameter hierarchy, was further broken down into three sub-patterns. Consistency was determined for both independent solution parameters (i.e. aircraft choice, resolution type, and vector direction) and combinations of these. Only three participants (1, 7 and 10) were found to have solved the designed conflict according to the most consistent solution parameter combination. Of these three, only participant 7 attained the highest level of consistency. For the other consistency groups (control problem analysis, solution geometry, and number of interactions) the highest consistency behaviour was determined for seven participants. Interesting is that participant 7 was assigned to each group.

Consistency group	Description	Pattern	Participants
Solution parameter hierarchy (1st interaction only)			

Aircraft choice	Consistently choosing either QS1338 or OM3185.	QS1338 OM3185	3, 5*, 7*, 13 1*, 10,
Resolution type	Consistently choosing the same resolution type	Heading	1*,3*,4*,6*,7*,8,9*, 10*,11*,12*,13,14*
Vector direction	Consistently choosing a left or right vector.	Left Right	14 1, 4, 7*, 10, 8*, 13**, 2***
Aircraft choice & resolution type		QS1338, heading OM3185, heading	3, 7* 1*, 10
Aircraft choice & resolution type, & vector preference		QS1338, heading right OM3185 heading right	7* 1, 10
Control problem pattern	Controlled aircraft in front or behind intruder aircraft	In front Behind	3, 7*, 8, 13* 1, 4, 9*, 10*, 11*, 12*, 14
Solution geometry	The same aircraft consistently taken behind the other.	OM3185 behind QS1338 QS1338 behind OM3185	1, 4, 2*, 7*, 8 10, 13 14
Number of interactions	Either one or two interactions were made to solve the conflict.	One interaction Two interactions	1, 9, 10, 14 3, 4, 6, 7*

* These participants displayed the most consistent behaviour, with matching behaviour in all four scenario repetitions.
**For one scenario the solution consisted of a combined resolution type involving a right turn and speed increase.
*** For two scenarios the solution consisted of a combined resolution type involving a right turn and speed increase.

Answers recorded in the post-simulation questionnaire generally supported the findings derived from the simulation data. In one question, participants were provided with a screenshot of the designed conflict and asked how they would prefer to solve it. Four participants stated that they would vector QS1338 behind OM3185. No participants explicitly stated that they would vector OM3185 behind QS1338. This is noteworthy since vectoring OM3185 to the right, behind QS1338 was one of the most common implemented solutions in the simulation (34% of all solutions). Investigating participant explanations for their solutions showed that vectoring one aircraft behind the other is preferred because of safety, efficiency and comfort. One participant stated: *"Easier to go behind an other aircraft then try to fly in front of it. It feels safer and is easier if you have to adjust the heading later on."* Participants who stated that they would vector both aircraft often motivated their solution by it being a less intrusive interaction and more fair treatment. One participant made the following argument: "*Divide the delay across the two a/c."*

4 DISCUSSION

In two separate real-time simulations, involving trainees and experienced controllers, we investigated operator consistency and consensus in solving CD&R tasks. For overall scenario performance, participants were found to be quite homogenous. The number of SSD inspections, interventions, and the distribution between speed, heading and combined interventions did not vary notably between participants. This result contrasts previous findings in the MUFASA project [7]. Here, controllers were found to differ in their overall scenario performance, with clear individual resolution type styles. Some preferred to work more with speed and combined control actions over heading. Naturally, the results obtained here are difficult to compare with those attained in the MUFASA project as scenario(s) and designed conflicts were different.

4.1 Consistency patterns

Participants varied considerably in how consistently they solved the designed conflict. Four different groups of consistency patterns were identified. These were based on:

- Solution parameters hierarchy analysis;
- Number of interactions undertaken to solve the designed conflict;
- Solution geometry; and
- Control problem analysis.

The further down consistency can be established in the solution parameter hierarchy, the stronger consistency. Of all participants, only three reached the third stage (aircraft match, resolution type, and direction). In isolation, however, the solution parameters hierarchy is less strong in terms of consistency. There is a difficulty in determining a threshold for consistency at these levels. Especially since all three involve only two options (choice of aircraft A or B, resolution type being heading or combination, and vector direction left or right). In combination, however, the finding of a consistent behaviour is more relevant. While the solution parameter hierarchy provides a rational method for analysing conflict solution consistency, a weakness of this approach is that it does not consider the spatial relationship, or traffic pattern, between the aircraft in conflict.

The spatial aspects pertaining to the solution were better captured by the analysis of solution geometry and control problem relationship. In combination, the results of the control problem and solution geometry analysis suggest that participants agreed on having aircraft B going behind A, but disagreed on how to achieve this traffic pattern. This suggests that consensus can be reached at a higher level, in terms of the desired solution geometry, but that there is disagreement on how implement this solution. It is not clear whether this discrepancy is relevant, or whether participants more arbitrarily chose which control actions to undertake in order to create the desired solution geometry.

4.2 Defining consistent behaviour

Allocation of a participant's behaviour to one consistency group did not preclude consistent behaviour as defined by any of the other groups. In fact, several participants were found consistent in as many as three groups. A few were even found to be consistent in all groups. Three participants were especially interesting since they were found to be consistent in contrast to each other. Two of these participants consistently (in three out of four scenarios) solved the designed conflict by a single interaction, vectoring aircraft B to the right behind A. In contrast, another participant consistently interacted twice, vectoring aircraft A to the right in front of B. Note, however, in terms of solution geometry, these seemingly contrasting solutions can be considered equal in that aircraft A always went in front of B. The difference lies in how participants achieved this solution geometry. Depending on how the conflict solution was analysed, participants could be determined to both match and diverge in their solutions. This finding was contradictory and surprising. It can, however, explain why previous studies have found controllers to be homogenous (e.g. [4-5, 8]). That is, at higher levels, controllers may be able to agree on how to solve conflicts. But when it comes to specific control actions, controllers may disagree.

4.3 Differences between trainees and experienced controllers

We also looked at differences between trainees and experienced controllers. With fewer SSD inspections, but more interactions, experienced controllers appeared to be more efficient than trainees. Furthermore, controllers perceived scenarios less difficult than did trainees. The choice of aircraft to intervene first with was evenly distributed both in the trainees group and experienced controllers group. This can be interpreted in two ways. Participants may have been very selective in their choice of aircraft to intervene with. However, it is possible that they did not consider the choice of aircraft important and that the observed distribution is the result of a random selection. The data for the choice of resolution type and vector direction can be interpreted in a similar way. This being said, the consistency analysis suggests the following: Participants performance did not indicate consensus on which aircraft to interact with first. It is, however, unclear how relevant this choice is and whether it would cause disagreement between controllers in an operational context. Regardless of the aircraft chosen, the data suggests that trainees agreed on right vectors being the best conflict solution. Among experienced controllers, there was a more even distribution between left and right vectors. Both used heading as the primary resolution type to solve the conflict. It is not surprising that heading changes was the most common resolution type.

The control problem analysis, however, suggests a slightly more consistent behaviour among experienced controllers than trainees. Whereas all controllers could be assigned a consistent behaviour according to this classification, trainees could not.

4.4 Implications for ATM and automation design

For the operational area, as well as automation designers, the assumption of homogeneity facilitates the development of a “one size” system. That is, for our cognition there is only one size available. It is, however, more likely that there are in fact individual differences between controllers that must be recognised. In order to determine the prerequisites for future automation design, we need to determine the degree of both consistency and consensus. One of the most difficult problems may be to identify time thresholds for decision-aiding automation – when should a decision-aid provide an advisory? Perhaps ideally, a decision aid would provide support “just in time” when the operator needs it. Support provided too late would be redundant, as the problem already has been solved. Support provided too early may be inappropriate and interrupt the operator. This, however, may be desired in terms of decision-aids functioning as a “safety net.”

Personalised and individualised automation in ATM, however, is a debatable topic. With safety as a priority, there is a need for control and continuous measurement and evaluation of established safety measures. Control is achieved by standardised operational procedures. This is helpful in maintaining a shared picture between controllers, and facilitating handovers during work shifts. It allows for global performance and capacity constraints to be set. With more individualised operations, and automation, transfers between controllers, and supervising each other, may be more complicated in that people work differently. Performance and capacity levels might vary between individuals, and affect sectorisation and staffing. Similarly, safety measures and key performance metrics may vary more between controllers. These risks will require further consideration.

5 CONCLUSION

In solving the designed conflict, we observed a variation in consistency patterns between participants. Conflict solution analysis revealed that participants differed in how consistent they were. Furthermore, patterns of consistency varied between participants, indicating that conflict solution preferences vary between individuals. Consistency, however, is highly dependent on the definition of what constitutes consistent behaviour. For reasons of automating, and other reasons such as possible operational impacts, we might have to consider differences in consistent behaviour within and between controllers. Furthermore, controllers were generally not found to be consistent in terms of when to interact. Time of interaction, however, is very important for the options available to solve a conflict. For design of decision-aid automation, large individual variations in interaction time may complicate calibration of suitable timing.

ACKNOWLEDGEMENTS

This work was supported under the SESAR Innovative Research Programme (Workpackage E), and by EUROCONTROL. We would like to thank LFV, and especially the staff at Norrköping TMC and the training department at Malmö ATCC, Sweden, for their support in hosting, facilitating, and participating in the simulations.

REFERENCES

[1] Klien, G., Woods, D. D, Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *Intelligent Systems, IEEE, 19*(6), 91-95.
[2] Kozhevnikov, M. (2007). Cognitive styles in the context of modern psychology: Toward an integrated framework of cognitive style. *Psychological Bulletin, 133*(3), 464-481.
[3] Cramer, H., Evers, V., Ramlal, S., Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction, 18*(5), 455-496.
[4] Kirwan, B., & Flynn, M. (2002). Towards a cognitive tool for conflict resolution assistance: A literature review. Rep. ASA.01.CORA.2.DEL04-A.LIT. Brussels, Belgium: EUROCONTROL
[5] Mogford, R. H., Guttman, J. A., Morrow, S. L., & Kopardekar, P. (1995). The complexity construct in air traffic control: A review and synthesis of the literature DOT/FAA/CT-TN95/22. Springfield, VA: National Technical Information Service
[6] Flynn, G. M., Leleu, C., & Hilburn, B. (2005). A complexity study of the Maastricht Upper Airspace Centre. EEC Rep. 2006-003. Brussels, Belgium: EUROCONTROL
[7] Hilburn, B., Westin, C., & Borst, C.. (2014). Will controllers accept a machine that thinks like they think? The role of strategic conformance in decision aiding automation. *Air Traffic Control Quarterly, 22*(2), 115-136.
[8] Magyarits, S., & Kopardekar, P. (2001). Dynamic density: The measure of air traffic control sector complexity for the en route environment. Phase II Experiment Plan. Rep. DD1-25-01. William J. Hughes Technical Center, FAA. http://www.tc.faa.gov/acb300/Techreports/DD_testplan.pdf (Accessed 13 July 2015).
[9] Westin, C. (2012). *Investigating controller reliability in air traffic complexity judgments.* Master's thesis. Nottingham University, UK
[10] Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide whether someone is an expert or not. *European Journal of Operational Research, 136*(2), 253-263
[11] Thomas, R. P., Willem, B., Shanteau, J., Raacke, J., & Friel, B. M. (2001). *Measuring performance of expert controllers: An application to air traffic control*. In proc. HFES 45th Annual Meeting, Minneapolis/St. Paul, MN, USA.
[12] Jha, P. D., Bisantz, A. M., Parasuraman, R., & Drury, C. G. (2011). Air traffic controllers' performance in advance air traffic management system: Part I — performance results. *The International Journal of Aviation Psychology, 21*(3), 283-305
[13] Mercado Velasco, G.A., Mulder, M., & van Paassen, M.M. (2010). *Analysis of Air Traffic Controller workload reduction based on the solution space for the merging task*. In proc. AIAA Guidance, Navigation, and Control Conference, Toronto, Canada.