TUDelft

Delft University of Technology

Preventing Scenario Recognition in Human-in-the-Loop Air Traffic Control Research

de Rooij, G.; Borst, C.; van Paassen, M.M.; Mulder, Max

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

PREVENTING SCENARIO RECOGNITION IN HUMAN-IN-THE-LOOP AIR TRAFFIC CONTROL RESEARCH

Gijs de Rooij, Clark Borst, M. M. (René) van Paassen and Max Mulder
Aerospace Engineering - Delft University of Technology, Delft, The Netherlands

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

Introduction

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

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

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

Mitigation Techniques

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

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

Study	Geometric		Textual		Temporal	
	Rotation	Mirroring	Renaming callsigns	Renaming waypoints	Time shifting	Reordering
Abdul Rahman, 2014	✓	-	-	-	-	-
Albuquerque et al., 2008	-	?	-	-	✓	✓
Borst et al., 2017	✓	-	-	-	-	✓
Borst et al., 2019	✓	✓	-	-	-	-
Cummings et al., 2005	✓	-	-	-	-	-
Harrison et al., 2014	-	-	✓	-	-	-
Hilburn et al., 2014	✓	-	-	✓	-	-
IJtsma et al., 2022	✓	-	-	-	-	-
Jans et al., 2019	✓	-	-	-	-	-
Jasek et al., 1995	-	-	-	-	-	✓
Jha et al., 2011	✓	-	✓	✓	-	-
Kim et al., 2022	✓	-	✓	✓	-	-
Klomp et al., 2016	✓	-	✓	-	-	-
Major and Hansman, 2004	✓	✓	-	-	-	-
Metzger and Parasuraman, 2006	✓	-	-	-	-	-
Rovira and Parasuraman, 2010	✓	-	✓	✓	-	-
Sollenberger and Hale, 2011	-	-	✓	-	-	-
ten Brink et al., 2019	✓	-	-	-	-	-
Trapsilawati et al., 2021	✓	-	✓	✓	✓	-
Wilson and Fleming, 2002	-	-	✓	-	-	-

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

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

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

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

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

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

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

Data Description

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

Participants and Apparatus

Ten professional en-route ATCOs (age: $\mu = 42.7$, $\sigma = 6.8$, years of experience: $\mu = 18.8$, $\sigma = 6.2$), from Eurocontrol's Maastricht Upper Area Control Centre (MUAC) voluntarily participated in a simulator experiment, as approved by the Human Research Ethics Committee of TU Delft under number 2754. All participants provided written informed consent. A TU Delft-built medium-fidelity simulator was designed to mimic the MUAC interface on a 1920 x 1920 pixels 27" display with a computer mouse for control inputs, shown in Fig. 1. Although the scenarios were static, participants could measure predicted minimum separation between flights and display extended flight labels.

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

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

Scenarios

Participants assessed five distinct static scenarios, intermingled with ten dummy scenarios that were not included in the current analysis. Each scenario was shown three times with different transformations and featured an artificial, octagonal 80 x 80 NM sector, with four waypoints in the cardinal directions. This made sure that the professionals would not fully rely on their trained scan patterns and that scenarios could not be recognized based on the sector shape. Four flights were present on direct routes to their exit points. In the dummy scenarios there were only two or three flights, for which measures like fixation orders would be less robust. Variants were created by applying any (combination) of the following transformations:

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

Callsigns were randomized for all variants and flight labels were always placed at a 90 degree offset to the direction of travel. Figure 2 shows an example of a scenario with corresponding transformations. Note that flights in the center of the sector were invariant to all geometric transformations and always appeared at the same location on the screen.
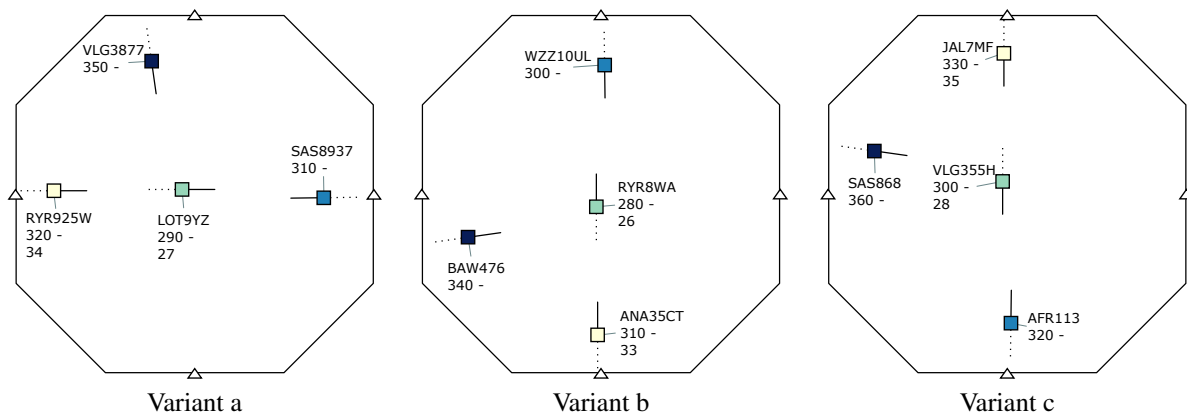
Figure 2: Three transformations of Scenario 5. Colors relate to the same flights in each transformation.

All participants got to see the same order of transformations, but the scenario ordering was counterbalanced between them to account for learning effects. This ordering of scenarios was defined in the previously executed experiment and might, in hindsight, have been suboptimal for the current study. The experiment started with six training scenarios.

Participant Task

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

Results and Discussion

After the experiment, some participants mentioned that they did recognize the repetition of certain conflict geometries, but none of them recalled that it were identical scenarios apart from the applied transformation(s). Our present analysis stays away from concluding whether the recognition mitigation has worked and instead focuses on the consistency of fixation behavior. Since participants showed vastly individualized behavior, no between-participant comparisons are performed and all observations discussed here relate to the three scenario repetitions per individual.

Fixation Order

Conflict detection time is directly driven by the order in which flights receive attention, especially when scenarios include many flights. After all, if an ATCO fixates flights in a different order, he/she might observe a conflicting pair earlier or later. To this end, Fig. 3 shows for each scenario's three repetitions the flight that was first fixated by each ATCO. The level of consistency, in terms of identical first fixations for all three transformations (visible as a row of three similarly colored squares), varied per ATCO from zero (Participants 5 and 10) to three scenarios (Participants 7 and 8). A similar variance can be seen between the scenarios, with consistent first fixations for one (Scenarios 1 and 2) to five (Scenario 5) ATCOs. This suggests that the rotations may have had an impact on the fixation order, and that this can differ per individual and traffic layout. On closer inspection, in 80% of the runs, the first fixated flight in Scenario 5 was located in the center of the sector (and therefore in the exact same location for all repetitions). Conversely, Scenario 1, the only one with no flight near the center, shows the lowest level of consistency.

To illustrate individual differences, complete orders of fixation for two participants on either extremes of the aforementioned consistency scale are shown in Fig. 4. Note how Participant 8's complete fixation sequence is consistent for all variants of Scenario 3. This, in combination with the inconsistent fixation orders seen in other scenarios or with other participants, further hints at a non-negligible influence of scenario rotation on the processing of traffic scenarios. For more insight in the relevant mechanisms, an analysis of scan patterns at different transformations would be useful, but this requires scenarios with more flights. The static, low density scenarios used in this study imply that the results are not necessarily applicable to dynamic and/or denser scenarios.
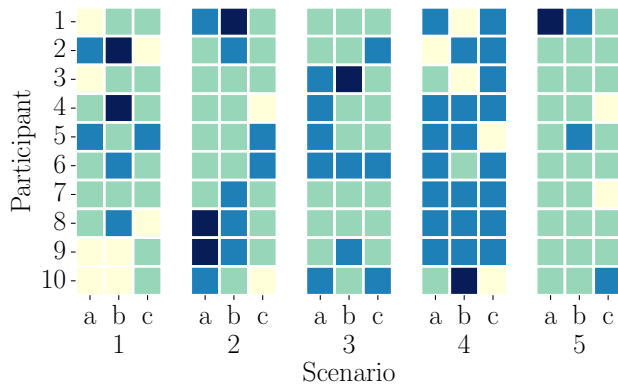
Figure 3: First fixated flight per participant. Colors represent specific flights in a scenario (see Fig. 2 for Scenario 5).
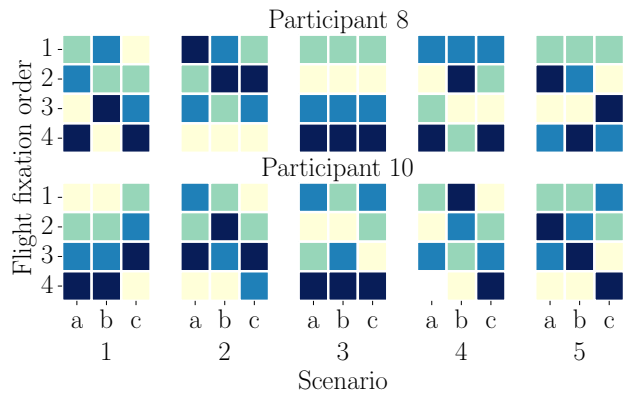

Figure 4: Complete flight fixation orders of two participants. Colors represent specific flights in a scenario (see Fig. 2 for Scenario 5).

Fixation Speed

To further illustrate the potential influence of rotations on fixation sequences and duration, Fig. 5 shows the standardized time till specific flights in Scenarios 3 and 5 had been first fixated. Results imply that the rotational-influence on this measure is dependent on the researcher's flight of interest. This is most visible in Scenario 5b, where Flight 1 shows significantly different means compared to the other two rotations. Akin to the fixation order, differences between individuals are again considerable, reflected in the wide spread of most data.
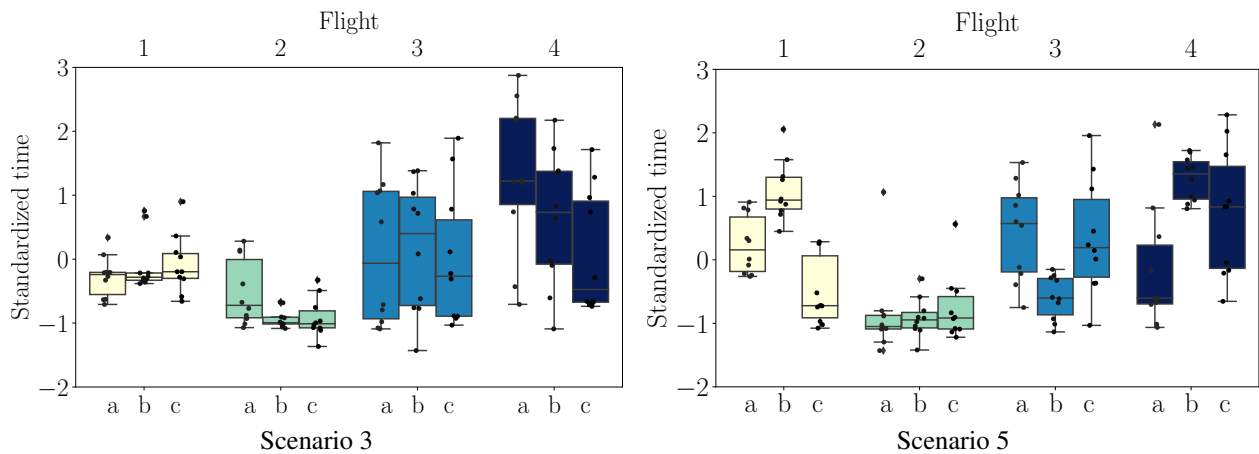

Figure 5: Standardized (per participant) time till flights have been first fixated in two scenarios, split per transformation. Colors represent specific flights in each scenario (see Fig. 2 for Scenario 5).

While the order of scenarios was counterbalanced between participants, the order of their repetitions was not (i.e., all participants first saw a, followed by b and then c). While this resulted in a clearly visible reduction in total fixation time over the three repetitions, this reduction is not (always) reflected in the results presented here. We therefore conclude that this speed-up was mostly caused by the participants getting more acquainted with the task at hand and advancing to the next scenario, rather recognizing the specific scenarios. To further isolate the effect of purely the transformation, future studies should include duplicate scenarios where no transformation has been applied.

Conclusion

Scenario transformations such as rotation and mirroring are proven techniques to create paired-samples in human-in-the-loop ATC research, but the potential impact on results is not always sufficiently recognized. We showed that the most popular technique, rotating scenarios, does risk eliciting different eye fixation behavior from participants,

potentially confounding objective measures such as conflict detection time. Whether this is a problem strongly depends on the research at hand and requires careful consideration. No definitive conclusions regarding the size of these effects can be made on the basis of the limited analysis presented here. The first indications do warrant further research with more elaborate, potentially dynamic, traffic scenarios and a tailored experiment design.

Acknowledgments

References

Abdul Rahman, S. (2014). *Solution Space-based Approach to Assess Sector Complexity in Air Traffic Control* (Doctoral dissertation). Delft University of Technology.

Albuquerque, E. A. F., Trabasso, L. G., Sandes, A., de Araújo, M., Li, L., & Hansman, R. J. (2008). Experimental Setup for Air Traffic Control Cognitive Complexity Analysis. *Symposium of Operational Applications in Areas of Defense*, 342–347.

Borst, C., Bijsterbosch, V. A., van Paassen, M. M., & Mulder, M. (2017). Ecological interface design: supporting fault diagnosis of automated advice in a supervisory air traffic control task. *Cognition, Technology and Work*, *19*(4), 545–560.

Borst, C., Visser, R. M., van Paassen, M. M., & Mulder, M. (2019). Exploring Short-Term Training Effects of Ecological Interfaces: A Case Study in Air Traffic Control. *IEEE Transactions on Human-Machine Systems*, *49*(6), 623–632.

Cummings, M. L., Tsonis, C. G., & Cunha, D. C. (2005). Complexity Mitigation Through Airspace Structure. *International Symposium on Aviation Psychology*, 159–163.

Fraga, R. P., Kang, Z., Crutchfield, J. M., & Mandal, S. (2021). Visual Search and Conflict Mitigation Strategies Used by Expert en Route Air Traffic Controllers. *Aerospace*, *8*(7).

Harrison, J., Izzetoğlu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., Woo, H., Shewokis, P. A., Bunce, S. C., & Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. *IEEE Transactions on Human-Machine Systems*, *44*(4), 429–440.

Hessels, R. S., Niehorster, D. C., Kemner, C., & Hooge, I. T. (2017). Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, *49*(5), 1802–1823.

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.

IJtsma, M., Borst, C., van Paassen, M. M., & Mulder, M. (2022). Evaluation of a Decision-Based Invocation Strategy for Adaptive Support for Air Traffic Control. *IEEE Transactions on Human-Machine Systems*, *52*(6), 1135–1146.

Jans, M., Borst, C., van Paassen, M. M., & Mulder, M. (2019). Effect of ATC Automation Transparency on Acceptance of Resolution Advisories. *IFAC PapersOnLine*, *52*(19), 353–358.

Jasek, M., Pioch, N., & Zeltzer, D. (1995). Enhanced Visual Displays for Air Traffic Control Collision Prediction. *IFAC Proceedings Volumes*, *28*(15), 553–558.

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. *International Journal of Aviation Psychology*, *21*(3), 283–305.

Kassner, M., Patera, W., & Bulling, A. (2014). Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1151–1160.

Kim, M., Borst, C., & Mulder, M. (2022). Situation Awareness Prompts: Bridging the Gap between Supervisory and Manual Air Traffic Control. *IFAC-PapersOnLine*, *55*(29), 13–18.

Klomp, R. E., Borst, C., van Paassen, M. M., & Mulder, M. (2016). Expertise Level, Control Strategies, and Robustness in Future Air Traffic Control Decision Aiding. *IEEE Transactions on Human-Machine Systems*, *46*(2), 255–266.

Major, L. M., & Hansman, R. J. (2004). *Human-Centered Systems Analysis of Mixed Equipage in Oceanic Air Traffic Control* (tech. rep.). MIT International Center for Air Transportation. Cambridge, MA, USA.

Metzger, U., & Parasuraman, R. (2006). Effects of automated conflict cuing and traffic density on air traffic controller performance and visual attention in a datalink environment. *International Journal of Aviation Psychology*, *16*(4), 343–362.

Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human Factors*, *52*(3), 411–425.

Sollenberger, R. L., & Hale, M. (2011). Human-in-the-Loop Investigation of Variable Separation Standards in the En Route Air Traffic Control Environment. *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*, 66–70.

ten Brink, D. S., Klomp, R. E., Borst, C., van Paassen, M. M., & Mulder, M. (2019). Flow-based air traffic control: Human-machine interface for steering a path-planning algorithm. *2019 IEEE International Conference on Systems, Man and Cybernetics, SMC 2019*, 3186–3191.

Trapsilawati, F., Chen, C. H., Wickens, C. D., & Qu, X. (2021). Integration of conflict resolution automation and vertical situation display for on-ground air traffic control operations. *Journal of Navigation*, *74*(3), 619–632.

Westin, C., Borst, C., & Hilburn, B. (2016). Strategic Conformance: Overcoming Acceptance Issues of Decision Aiding Automation? *IEEE Transactions on Human-Machine Systems*, *46*(1), 41–52.

Wilson, I. A. B., & Fleming, K. (2002). Controller reactions to free flight in a complex transition sector re-visited using ADS-B+. *Proceedings. The 21st Digital Avionics Systems Conference*, *1*, 5–1.