2013

Mismatches Between Automation and Human Strategies: an Investigation into Future Air Traffic Management Decision Aiding

Carl A. Westin

Clark Borst

Brian Hilburn

MISMATCHES BETWEEN AUTOMATION AND HUMAN STRATEGIES: AN INVESTIGATION INTO FUTURE AIR TRAFFIC MANAGEMENT DECISION AIDING

Carl A. Westin[1], Clark Borst[2] & Brian Hilburn[1]
[1]*Center for Human Performance Research (CHPR), Voorburg, The Netherlands*
[2]*Delft University of Technology, Delft, The Netherlands*

Future air traffic management will have to rely on more, and more sophisticated, automation to accommodate predicted air traffic. However, studies across various domains have shown that user acceptance of automation decreases when the authority of decision-making automation increases. As a result, low user acceptance could lead to disuse of an automated tool and threaten potential safety and performance benefits. Through a series of human-in-the-loop simulations, the work described in this paper examined the interacting effects of air traffic complexity and strategic conformance, i.e., the fit between human and machine strategies, on automation acceptance in a conflict detection and resolution task. An experiment with 16 professional air traffic controllers showed that strategic conformance is a potentially important construct. That is, conformal resolution advisories were more accepted, led to higher controller agreement, and also reduced response time to proposed advisories.

Future Air Traffic Management (ATM) will have to rely on more, and increasingly sophisticated, automation to accommodate predicted air traffic. This entails automation growth in terms of types of tasks it can perform, and the level of authority and autonomy it can assume, a prediction captured in the SESAR program's definition of five operational Service Levels (SESAR, 2007) intended to guide European ATM evolution.

Studies across various domains have shown that user acceptance of automation decreases when the authority of decision-making automation increases (Kaber & Endsley, 2004; Parasuraman & Riley, 1997). Research has also shown that the predominantly algorithmic approaches used in automation seldom fit well with the more heuristic processes employed by humans. For example, in their exhaustive survey of conflict detection and resolution (CD&R) modeling methods, Kuchar and Yang (2007) concluded that CD&R automation correlates poorly with how controllers prefer to work. Bekier, Molesworth, and Williamson (2012) suggested that there is a "tipping point" for any automated tool, above which controller acceptance of that tool quickly drops. Consequently, a low user acceptance of an automated tool may lead to disuse (and abuse) of that tool (Parasuraman & Riley, 1997) which could severely undermine the intended safety and performance benefits of the tool.

The work in this paper primarily examines the notion of *strategic conformance*, i.e., the extent to which automation's performance and underlying processes are similar to those of the human. Additionally, the interacting effects of strategic conformance with air traffic complexity will also be addressed. Whereas there has been a great deal of empirical and theoretical work into automation in relation to traffic complexity, much less has been done in the area of strategic conformance. Paradoxically, this could in the future become the most critical issue of all, as mismatches between human and machine could threaten initial acceptance of advanced automation. The method applied in this paper for studying human-machine conformance involves presenting controllers unrecognizable replays of either their own solution (i.e., a *conformal solution*) or a colleague's different solution (i.e., *non-conformal*) to a series of pending aircraft separation conflicts. Controllers were instructed that such solutions were generated by automation. Assuming controllers remain consistent in their solution choice over time, this allows us to experimentally manipulate conformance between human and (simulated) machine solutions. Note that by replaying human actions we are in fact simulating automation capable of providing conflict resolutions of the same caliber as a human controller would be able. This replay procedure is inspired by one used many years ago, with which it was shown that operators might be more likely to find faults in automation than in themselves, even when "automation" is an unrecognizable replay of their own performance (Fuld, Liu, & Wickens, 1987).

This paper intends to answer the following questions: are controllers more likely to accept automated advisories when the advisory mimics their own solution? Will acceptance of automation vary with air traffic complexity, regardless of solution conformance? Will workload increase with non-conformal resolution advisories? Does response time vary with conformance of the advisory? First the experimental design is described. The next section presents the results of the study, followed by a discussion and conclusion in the final section.

Experimental Design

We conducted a series of two human-in-the-loop Air Traffic Control (ATC) simulations of increasing automation authority. The first simulation series, considered to be a *prequel* to the main conformance experiment, was designed to capture controllers' manual performance in maintaining safe separations between aircraft. The controllers needed to use an advanced separation assistance tool, the Solution Space Diagram (SSD) developed by Delft University of Technology (Figure 1), to vector aircraft and solve conflicts by issuing speed and/or heading clearances to aircraft (Mercado-Velasco, Mulder & Van Paassen, 2010). In its most succinct form, the SSD is a tactical decision-support tool that visualizes 'go' and 'no-go' areas on a circular heading ring around an aircraft. When vectoring an aircraft into a 'go' area, it will remain free of conflict with all other aircraft and a loss of separation will essentially never occur. In the second simulation series, i.e., the conformance experiment, the controller solutions (and those of their colleagues) to specific conflicts were replayed as automation advisories and plotted within the SSD, after which the controller could either accept or reject the advisory. By plotting the advisory within the SSD controllers could always inspect the validity of the advisory, but also "look around" for better alternatives.

Subjects

Sixteen professional air traffic controllers voluntarily participated at the Shannon Area Control Centre, Ireland. Experience ranged from zero to ten years ($\bar{x}$ = 2.5 years). Twelve controllers currently worked en-route and one controller worked the tower position. Three were students currently undergoing en-route training.

Apparatus

The ATC simulator ran on a portable computer connected to an external 21" monitor. Participants interacted with the simulator through an externally connected computer mouse and keyboard. The ATC simulator was a Java-based application (using OpenGL extensions) that allowed air traffic controllers to control short traffic scenarios. To vector an aircraft, a controller used a computer mouse to click on an aircraft of interest, drag the velocity trend vector to a new conflict-free area on the heading ring (a "clear" area outside the red/yellow areas), and press the ENTER key on a keyboard to implement the vector. Speed clearances (and combined speed and heading clearances) could also be given by using the mouse scroll wheel to either increase or decrease speed. This also increased the radius of the heading band and showed the corrected conflict zones for the new speed settings. This allowed a controller to quickly browse through different speed settings and preview the conflict and conflict-free heading areas for different speeds. Further, no wind conditions were taken into account, all aircraft remained on the same flight level and could not be changed, and the aircraft velocities (and speed clearances) were given in knots Indicated Airspeed (IAS). Further, the aircraft motion was simulated by first order, linear kinematic equations and to keep traffic scenarios sufficiently short and interesting, it was decided to run the simulator four-times faster than real time. Speeding up traffic scenarios in ATC simulators is a common technique to serve this purpose. Finally, the aircraft plots on the display were updated every second to simulate a 1 Hz radar update frequency.

Traffic Scenarios

Each series consisted of 16 traffic scenarios, each based on a squared airspace equal in size (Figure 1). Four baseline scenarios were each rotated in different angles to create three variants, resulting in four scenario groups with four scenario variants in each group. This reduced potential confounding factors, and ensured that initial complexity was the same across scenarios, facilitating comparison between low and high complexity conditions. We aimed to make each traffic scenario repeatable, yet unrecognizable to participants. We maintained sector geometries through scenario rotations in which the relative trajectories and closure angles of aircraft were kept constant, but the entire sector was rotated, and sector entry/exit points renamed.

To guarantee exact replays of controller solutions, each baseline scenario featured only one designed conflict between two aircraft. As a result, any other conflict occurring in a scenario was the consequence of controller intervention. The geometry of the designed conflict was only varied between baseline scenarios. The conflict pair was initially aligned to the exit points and thus required no initial controller interaction. The other aircraft in the sector were considered "noise" aircraft to distract the controller from the conflict pair. Some of the noise aircraft were misaligned with their exit point and displayed in grey, whereas aligned noise aircraft were displayed in green such that the controller could immediately see which aircraft had not yet been cleared to their exit point.

In designing the scenarios it was very important that noise aircraft not interfere with the designed conflict, such that in a replay scenario (i.e., the conformal experiment) the controller would have the same set of solution as

Figure 1. The ATC simulator, showing: 1) the current performance score, time, and runs to go, 2) the SSD of the currently selected aircraft in magenta, 3) the SSD with an automated resolution advisory in an amber color (suggesting a speed decrease to resolve a future conflict), and 4) the advisory dialog window that the controllers could use to either accept or reject the advisory.

observed during the prequel experiment. In the prequel experiment each scenario lasted roughly two minutes. In the conformal experiment each scenario lasted less than one minute in order to reduce the likelihood of significantly changing the traffic situation and thus not being able to guarantee solution replays (i.e., automation advisories) that would solve the conflict. Further, we decided to introduce the automation advisory early in each scenario to prevent participants from solving the designed conflict proactively.

Independent Variables

The experimental followed a within-subjects design with two independent variables, solution conformance (conformal vs. non-conformal) and traffic complexity (low vs. high). A conformal advisory was qualified in terms of aircraft choice, clearance type (e.g., heading change only), and clearance direction (e.g., heading change to the left). A non-conformal solution therefore always featured a different aircraft choice and/or clearance type and direction. Further, non-conformal solutions were derived from solutions provided by other controllers. Complexity was varied through means of aircraft count, and calibrated in a series of test trials. Finally, presentation order of traffic complexity and solution conformance was balanced between participants and traffic scenarios using a Latin Square design.

Dependent Measures

The dependent measures of the conformance study focused on the acceptance of the advisory, the controller’s agreement with the advisory (measured on a zero to one-hundred scale), response time, and subjective workload ratings (measured on a zero to one-hundred scale). Response time was measured from presentation of the resolution advisory to pressing the ‘accept’ or ‘reject’ button.

Experiment Procedure

The whole experiment took four weeks in total. In the first week, the prequel experiment was conducted with the aim to capture controller resolutions to the designed conflicts. Following briefing and consent procedures, we conducted 16 training runs and 16 measurement runs. Participants were given two main tasks to be performed using the SSD, namely resolving conflicts and clearing aircraft to their intended exit points. A continuously updated performance score reflecting these two task parameters was included to keep participants focused and motivated, and more importantly to prevent scenario recognition and early detection of the designed conflict. To warn the controllers for short-term conflicts, an auditory alert was triggered and the aircraft involved in the conflict were displayed in red.

In the second and third week, prequel data were analyzed and a set of eight conformal (i.e., a replay of controller's own decision) and eight non-conformal (i.e., a replay of a colleague's different but workable) advisories was created for each individual participant. Finally the conformance experiment was conducted in week four. Following a simulator briefing the experiment started with eight training runs, followed by 16 measurement runs. The same scenarios as in the prequel were used but the order varied according to a latin square design. Participants performed the same task as in the prequel experiment, but now were assisted by a higher level of automation that would once in a while provide resolution advisories by proactively auto-select an aircraft in conflict.

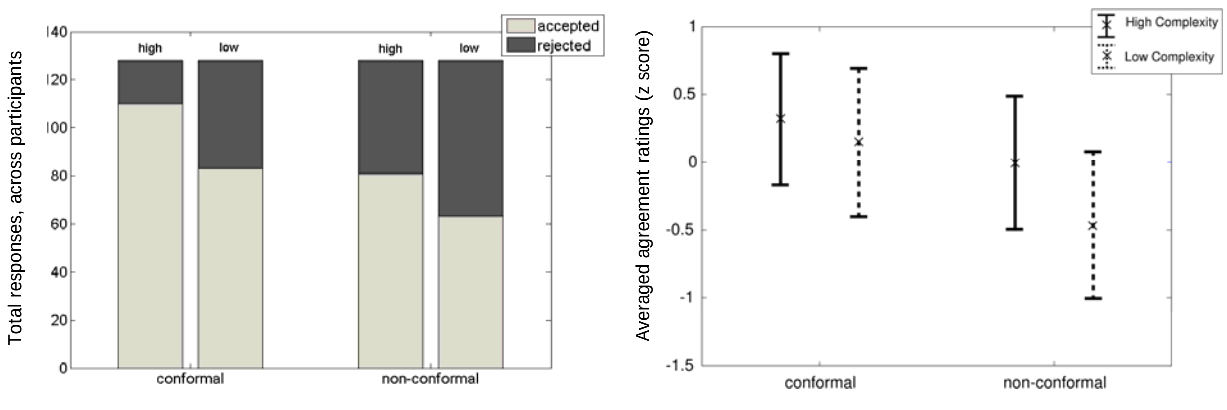
The resolution advisory consisted of either a heading vector, speed vector, or combination thereof, plotted inside the SSD of that aircraft. The resolution advisory was accompanied by a beeping sound and the appearance of a dialog window that the controller had to use to either 'accept' or 'reject' the advisory (see Figure 1). Upon pressing the accept button, the advisory would be automatically implemented. Upon pressing the reject button, the advisory was discarded and the controller had to implement his own workable solution using the SSD. Further, the controllers had 15 seconds to inspect, accept, or reject the advisory. After 15 seconds, the advisory expired and the controllers had to implement their own workable solution. Note that participants were told that an advisory would always solve the conflict, but it would not always suggest the most optimal solution. As such, controllers were encouraged to find better alternatives at their own discretion. After each scenario, participants were given performance feedback in terms of an average performance score. Second, controllers were asked to give ratings on their experienced workload and their agreement with the automation advisory. After the experiment, participants were asked to complete a questionnaire containing information of demographic value and statements (in five-point Likert scale format) querying participant's opinions of the simulator, the SSD interface, and the automated advisories.

Results

Experimenter observations and an analysis of questionnaire data indicated that participants did not recognize scenarios, and more importantly, that the automated advisories were, in half the cases, replays of their own prior conflict resolutions. Questionnaire data also showed that controllers enjoyed the simulator and SSD tool, but did not find scenarios very realistic.

Advisory Acceptance and Agreement with Resolution Advisory

Cumulative accept/reject scores for conformance and complexity can be seen in Figure 2(a). Participants accepted more resolutions advisories in the high complexity scenarios. Comparing conformance, it can be seen that participants accepted more conformal scenarios than non-conformal scenarios. A 2 x 2 repeated measures ANOVA showed a significant main effect for both complexity ($F(1,15) = 11.139$, $p<0.01$) and conformance ($F(1,15) = 10.624$, $p<0.01$) on acceptance. The interaction between complexity and conformance was not significant.



(a) *(b)*

Figure 2. Proportion of advisory acceptances (a) and error bar plots of normalized agreement ratings (b), by complexity and conformance.

Participant ratings of agreement with the resolution advisory revealed a difference between both the complexity levels and conformance levels (Figure 2(b)). The 2 x 2 repeated measures ANOVA revealed that both complexity ($F(1,15) = 7.735$, $p<0.05$) and conformance ($F(1,15) = 18.095$, $p<0.01$) had a significant main effect on the agreement ratings. Agreement with resolution advisories varied positively with increasing complexity and conformal scenarios.

Although a trend was observed in the interaction between complexity and conformance (Figure 3), this was not significant ($F(1,15) = 3.186$, $p>0.05$). This trend suggests that non-conformal solutions tended to be less agreed with than conformal solutions. This effect was especially apparent under low complexity, perhaps because controllers were under less time pressure and would have had more time to evaluate candidate solutions.

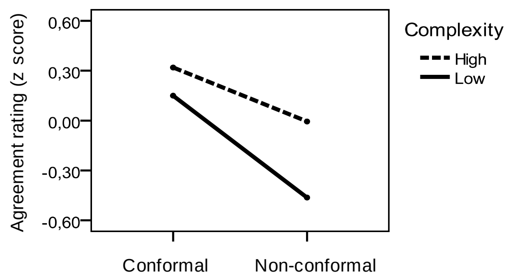


Figure 3. Agreement with resolution advisory ratings (normalized), by complexity and conformance.

Workload and Response Time

The 2 x 2 repeated measures ANOVA showed that workload significantly increased with more complex scenarios ($F(1,15) = 179.950$, $p<0.01$). Neither conformance ($F(1,15) = 0.397$, $p>0.05$), nor the interaction between the two factors was significant ($F(1,15) = 0.266$, $p>0.05$). For response time a main effect was measured by performing a 2 x 2 repeated measures ANOVA, with conformal scenarios having a significant faster response time ($F(1,15) = 9{,}557$, $p<0.01$). Although showing a trend, scenario complexity did not have an effect on response time ($F(1,15) = 4.182$, $p>0.05$). The interaction between complexity and conformance was also not significant.

Discussion

The main objective of this study was to investigate the effect of strategic conformance and complexity on controller performance and acceptance in the context of higher levels of automation. A significant main effect of conformance was observed on acceptance, agreement with resolution advisory, and response time. Conformal resolution advisories were accepted more often, received higher agreement ratings, and were acted upon faster than were non-conformal advisories. Results could, however, not confirm the hypothesis that conformal advisories reduced workload participant workload. These findings suggest that controllers not only discriminate between resolution advisories, but more importantly, prefer resolution advisories that match their own way of working in comparison to advisories generated by their colleagues. That response time increased for non-conformal resolution advisories suggest that controllers questioned the advisories before making the decision to either accept or reject, and that it took longer to conclude that the advisory did not fit with the controllers' own solution. These results support further research initiatives into heuristic-based automation.

Similar to conformance, complexity significantly affected acceptance and agreement with resolution advisory. In high complexity scenarios, acceptance and agreement with resolution advisory was higher than in low complexity scenarios. The lower acceptance and agreement observed in low complexity scenarios could be explained in that controllers had more time to question the advisory and consider more optimal options. The results with low complexity scenarios generating significantly lower workload ratings in comparison with high complexity scenarios, supports this. With increasing workload there are less resources and time available to consider options before making a decision. But these results are also interesting in that they show that controllers disagree with themselves when complexity is lower. It could be indications of learning effects, but it is also a potential indicative that controllers are inconsistent over time, or traits of automation bias.

In this study, controllers were told that all resolution advisories were generated by the automation. How would controllers react if the presumed source was varied between themselves, colleagues, and automation? Further research will attempt to better clarify the concept of automation bias by investigating the effect the presumed source of the advisory would have on acceptance.

It is interesting to see that response time varied with conformance but not complexity, and that workload varied with complexity but not conformance. It could be argued that workload and response time would somehow be connected. However, it is important to underline that response time is applicable to the designed conflict and resolution advisory only, whereas workload considers the entire scenario. The specific workload pertaining to the designed conflict and resolution advisory was not pursued.

The unique experimental design required careful consideration in designing dynamic traffic scenarios allowing exploration of strategic conformance. In order to contain the extent of confounding variables, and increasing experimental control we sought to limit the number of scenarios and through variations of scenario rotation angles increase the measurement span. This experiment showed that it is possible to create "fake" resolution advisories based on participant's own solutions, without participants later recognizing it as their own. It opens up new possibilities in researching not only conformance, but also other aspects such as within-participant consistency and reliability over time.

Despite our best efforts, questionnaire data revealed that controllers did not find scenarios very realistic. This was, however, not surprising considering the limitations of the simulator (i.e., no flight level changes, no wind, limited information, etc.) and that participants were instantly thrown into situations of various complexities. Although we have found some effects of complexity in our study, the results cannot simply be extrapolated to a real operational setting.

Conclusion

These results suggest that strategic conformance is a potentially important construct underlying automation acceptance, and that it can benefit both agreement with automation and response time. A trend toward a conformance and complexity interaction also suggests that the effect on controller agreement with automation advisories is more pronounced under low complexity situations. Though this experiment was considered an important first step, there are several remaining questions to be answered, with respect to both the definition of inherent automation bias, and the nature of such bias as it relates to presumed source of strategic advice.

Acknowledgements

The authors acknowledge the inspiration from EUROCONTROL and the SESAR Joint Undertaking. The work was co-financed by EUROCONTROL on behalf of the SESAR Joint Undertaking in the context of SESAR Work Package E (project MUFASA: Multidimensional Framework for Advanced SESAR Automation). This work reflects only the authors' views and EUROCONTROL is not liable for any use that may be made of the information contained herein. Finally, we would like to thank the staff at the Shannon Area Control Centre, Ireland, for all their amazing support in hosting, facilitating, and participating in the experiment.

References

Bekier, M., Molesworth, B.R.C., & Williamson, A.M. (2012). Tipping point: The narrow path between automation acceptance and rejection in air traffic management. *Safety Science*, 50(2):259-265.

Fuld, R. B., Liu, Y., & Wickens, C.D. (1987). The impact of automation on error detection: some results from a visual discrimination task. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 31(2):156-160.

Kaber, D.B. & Endsley, M.R (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.

Kuchar, J.K. & Yang, L.C. (2000). A review of conflict detection and resolution modeling methods. *IEEE Transactions on Intelligent Transportation Systems*, 1(4):179-189.

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. *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Toronto, Canada.

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

SESAR (2007). Deliverable D3: the ATM target concept, DLM 0612 001 0200a, September 2007.