Evaluation of a Decision-Based Invocation Strategy for Adaptive Support for Air Traffic Control

IJtsma, Martijn; Borst, Clark; van Paassen, Marinus M.; Mulder, Max

DOI
10.1109/THMS.2022.3208817

Publication date
2022

Document Version
Final published version

Published in
IEEE Transactions on Human-Machine Systems

Citation (APA)
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. https://doi.org/10.1109/THMS.2022.3208817

Evaluation of a Decision-Based Invocation Strategy for Adaptive Support for Air Traffic Control

Martijn IJtsma, Clark Borst, Marinus M. van Paassen, *Senior Member, IEEE*, and Max Mulder, *Senior Member, IEEE*

***Abstract*—Air traffic controller workload is a limiting factor in the current air traffic management system. Adaptive support systems have the potential to balance controller workload and gain acceptance as they provide support during times of need. Challenges in the design of adaptive support systems are to decide when and how to trigger support. The goal of this study is to gain empirical insights into these challenges through a human-in-the-loop experiment, featuring a simplified air traffic control environment in which a novel triggering mechanism uses the quality of the controller's decisions to determine when support is needed. The designed system seeks to prevent high workload conditions by providing resolution advisories when the controller exceeds a threshold of "self-complicating" decisions. Results indicate that the new system is indeed capable of increasing the efficiency and safety compared to full manual control without intervention. More adaptive support, however, increased the frustration of participants, decreased acceptance, and did not result in improved workload ratings. These findings suggest that, unless we can better infer human intent in complex work environments, adaptive support at the level of decision-making is problematic. A potentially more fruitful direction is to provide support at the level of information integration, with full decision-making authority with the human.**

***Index Terms*—Adaptive intervention system, air traffic control (ATC), decision quality, triggering threshold.**

I. Introduction

EXPECTED increases in air traffic volumes and operational complexity require the air traffic control (ATC) community to exploit more advanced automation support in order to maintain safe and orderly traffic flows at acceptable workload levels. Arguably one of the most difficult hurdles to overcome is ensuring acceptance of automation that aims to take over part(s) of the controller's tasks [2], [3]. The concepts underlying adaptive support systems, introduced by Rouse in 1976 [4], could provide a solution to both the workload and acceptance problems. Adaptive systems are the "*technological components of a joint human–machine system that can change their behavior to meet the changing needs of their users*" [5, p. 1008]. One class of adaptive systems is based on dynamic trading of control authority between the user and the adaptive support system, with the aim of off-loading the human user during periods of high workload. Providing support only when needed most could mitigate workload peaks. Designing an adaptive system that provides the right support at the right time, however, is challenging, with questions around what parameters would determine a "time of need," when the automation should intervene, and when it should give back control [5], [6], [7]. Research on adaptive systems has explored these questions in a variety of domains, such as naval combat [8], smart manufacturing [9], intelligence, surveillance & reconnaissance operations [7], and intelligent tutoring systems (ITS) [10], [11]. Design of adaptive systems requires decisions to be made about the following two interrelated factors:

1) the type of adaptation, which can be based on changing function allocation, task scheduling, interaction, and content;
2) the mechanism for triggering for adaptation, where existing mechanisms can broadly be classified as operator-based, system-based, environment-based, task- and mission-based, or spatio-temporal [5].

While many studies investigated the effectiveness of various of these triggering mechanisms, examples are studies examining psychophysiological measures [12], there is little empirically supported guidance for, once a triggering mechanism is decided, determining a threshold for invoking adaptation. Only a few studies of adaptive systems mention how a triggering threshold is set, e.g., by comparing profiles of the triggering variable and comparing them to known periods of high subjective workload [7] or through baseline experiments [13]. Addressing this gap is important because the threshold determines to a large degree the frequency of adaptation, with a lower triggering threshold resulting in more frequent adaptations, which in turn affects the operator's workload, acceptance, and the performance of the human–machine system. Without a good understanding of the implications of various triggering thresholds, the design of adaptive systems risks setting thresholds arbitrarily, reducing their effectiveness.

This article aims to investigate the effect of various triggering thresholds on operator workload, automation acceptance, and

Manuscript received 20 September 2021; revised 29 May 2022 and 4 August 2022; accepted 5 September 2022. Date of publication 25 October 2022; date of current version 15 November 2022. This work was presented in part at the International Symposium on Aviation Psychology, Dayton, OH, USA, May 811, 2017. This article was recommended by Associate Editor R. A. Hess. *(Corresponding author: Martijn IJtsma.)*

Martijn IJtsma is with the Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: ijtsma.1@osu.edu).

Clark Borst, Marinus M. van Paassen, and Max Mulder are with the Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands (e-mail: c.borst@tudelft.nl; m.m.vanpaassen@tudelft.nl; m.mulder@tudelft.nl).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by TU Delft Ethics Board.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2022.3208817.

Digital Object Identifier 10.1109/THMS.2022.3208817

system performance, providing an empirical exploration of the tradeoffs associated with setting a threshold. In addition, the study examines the potential effectiveness of an operator-based triggering mechanism, designed to prevent future workload bottlenecks and supporting workload-mitigating strategies in ATC. The majority of adaptive systems for mitigating transient workload peaks use triggers that are based on real-time measurement or estimation of current workload status, with no projection toward future states and task demands. Thus, the system adapts *after* an increase in workload has occurred (and has been detected). In work domains with a certain degree of predictability, a solely reactive approach can be problematic as it misses opportunities for anticipating future task demands and supporting an operator proactively to prevent workload bottlenecks. In ATC, evidence from naturalistic studies with expert controllers [14], [15] suggests that more proactive support can help reduce future task demands and workload.

By means of a human-in-the-loop (HITL) experiment, the study aims to gain empirical insights into the issue of threshold setting for adaptive systems, as well as to test a triggering mechanism for preventative workload management. This article first reviews related work on automation for ATC, triggering thresholds for adaptive systems, and ATC strategies. The design of the novel adaptive system is discussed in detail, including the mechanism for monitoring operator performance and the type of support that is invoked when the trigger variable exceeds a critical threshold.

II. Background

A. Automation Support for Conflict Detection and Resolution (CD&R)

Maintaining separation between aircraft is a core responsibility of ATC, with human controllers playing an essential role because of their ability to integrate information and make judgments in this complex work domain [16]. Tasks performed by air traffic controllers include CD&R. Designing advanced forms of automation to support air traffic controllers in CD&R is common [16].

The body of literature on CD&R automation is too large to discuss here but recent examples include the development of controller assistance tools (CATO) [17] and conflict resolution tools for flight-centric ATC [18], [19]. These tools provide automation support at the level of information acquisition and integration. Both studies have shown favorable results for reducing the controller workload and increasing airspace capacity.

Automation support at the level of action selection and implementation is typically more challenging, because it requires the controller to share control authority with the automation. One of the challenges with providing automation support at this level is its acceptance [2], [3]. A lack of acceptance might point to underlying issues with human–automation interaction, for example, the automation's lack of support for individual strategies or preferences [2]. Lack of acceptance can also be caused by automation being clumsy, or interfering with a controller's workflow when support is not needed or appreciated.

Adaptive systems, which change their behavior to fit the current context and user needs [5], could provide a solution to these challenges. An effective adaptive system relies on an unobtrusive assessment and interpretation of context—such as controller state, task demands, or performance—and, much like a human assistant, provides appropriate support in response to this context. Design of such an intervention system is anything but trivial [5], illustrated by the fact that research on adaptive systems has been primarily experimental, with few practical implementations.

B. Triggering Thresholds in Adaptive Systems

Design of adaptive systems requires formalized logic for adaptation. Two interdependent design choices determine the behavior of an adaptive system: the measures used to change behavior, and the thresholds at which these changes are triggered. A brief overview of research related to these design decisions is provided here. It focuses on systems for mitigating transient workload peaks, an important application for adaptive systems in ATC and similar domains. For a full review of other applications of adaptive systems and related forms of triggering logic, see Feigh et al. [5].

Measures for triggering adaptive support aimed at mitigating workload are most commonly based on operator measurement and/or assessment of environmental state [5], including instantaneous experienced workload (most commonly through electroencephalograms (EEGs), electrocardiograms (ECGs), and/or electromygrams (EMGs) [10], [12]), performance as an indicator of workload [5], and external events that impact workload [20]. The majority of these triggering mechanisms invoke support once they detect or infer an increase in workload. Thus, they only provide support *after* workload has increased. Based on research discussed in the next subsection, measures and triggering mechanisms focused on preventing future workload might be a more effective approach, especially when there are significant risks of downstream or cascading effects of poor decision-making under high workload conditions. A parallel can be drawn between adaptive systems for mitigating workload peaks and intelligent tutoring systems (ITS), as another application of adaptive systems, which often purposely provide support early-on to prevent students from going too far along an incorrect solution path [21].

Once a triggering variable is selected, the behavior of the adaptive system is determined to a large degree by the thresholds for invoking automation. Discussions of adaptive systems often focus on the types of trigger, not the thresholds. There is little empirically supported knowledge on the relation between triggering threshold and an adaptive system's effectiveness. Some guidance can be inferred from existing research, with studies on automation cycles [22], [23], in which control authority for a tracking task is switched back and forth between manual and automated control based on fixed time intervals, which show that more frequent switching (corresponding to low thresholds) increases performance but also results in higher workload ratings. Likewise, the ITS community has looked at optimal timing of feedback, ranging from systems that provide immediate

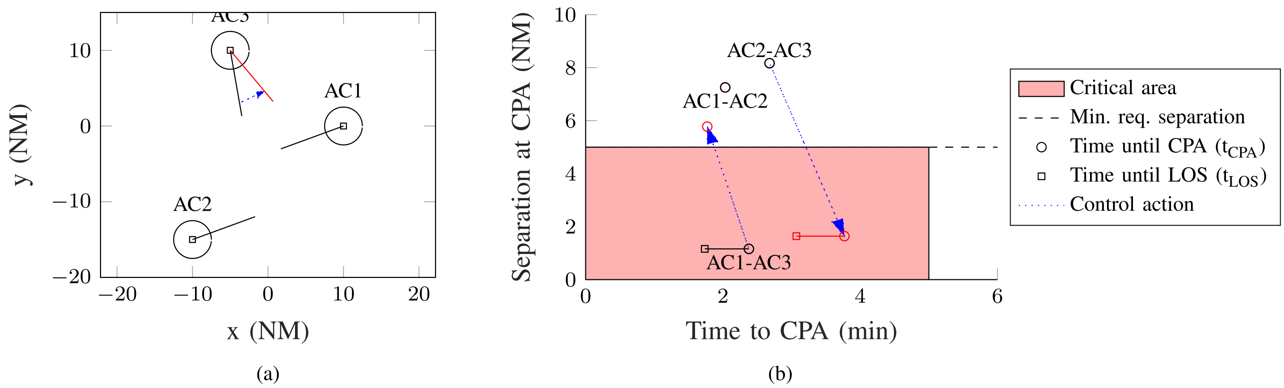

Fig. 1. Example of a separation monitor plot. (a) Traffic scenario. (b) Separation monitor plot.

feedback [21] to ones that withhold feedback until after the student has gone several steps along an incorrect solution path or after the student requests help [24]. Studies found that delayed feedback results in reduced time efficiency (i.e., a lower learning rate) compared to immediate feedback [21], [24], [25]. This suggests that a tradeoff exists between immediate intervention and a more "laissez-faire approach," balancing performance benefits of properly timed support with potential adverse effects of invoking/revoking support too often.

C. Preventing Future Workload

To analyze and design the interactions between a controller and an adaptive system requires understanding the cognitive strategies applied by controllers to manage their workload. Ethnographic studies of air traffic controllers [26], [27] show that controllers are effective in managing their own workload: Expert controllers switch their CD&R strategies in response to changing work demands and can thereby maintain performance under a wide range of complexity and workload conditions. For example, a "wait-and-see" strategy, in which controllers identify a conflict but then wait for the situation to evolve, requires significant monitoring effort and is, therefore, used during low workload conditions but not in high workload situations [26], [27]. Likewise, several of the best practices for conflict resolution are more economical in terms of the amount of monitoring that is required after implementation of the resolution [28], [29], [30]. This matches general patterns found in research on expert workers in complex work domains [31]: Experts switch to opportunistic but more cognitively economical strategies during times of high task demands and, vice versa, employ more strategic yet cognitively intensive strategies when task demands are low.

During high workload situations, however, controllers often inadvertently exacerbate task demands when making opportunistic or tactical decisions. One study found that for one out of two conflict resolutions following a short-term conflict alert (STCA), the controller induced downstream conflicts with additional STCAs [14]. Likewise, a similar experiment found that high workload and complexity conditions resulted in several participants creating additional conflicts, increasing workload even further [15]. In these instances, decision-making can be locally optimal, given the constraints on time and cognitive resources, but fails to consider the global traffic situation resulting in cascading effects that can further increase task demands.

A "self-complicating" decision is made when a controller chooses an action, typically as a tactical measure, that may temporarily solve a conflict, but creates a more complex traffic situation, with conflicts introduced by the decision. Triggering adaptive support based on increases in complexity can be seen as a type of performance-based measure [5], but rather than just indicating current high workload conditions, it also acts as a predictor of future, downstream workload. Evaluating the complexity of a future traffic situation is, however, not trivial, as it would require comparison of chosen decisions against a norm. As a practical measure, we count decisions as self-complicating when they create one or more new conflicts.

III. Adaptive System Design

Based on controllers' strategies, an effective adaptive system for CD&R should help prevent cascading effects and self-induced workload by providing support during high workload situations, and assist controllers in managing their own future workload by promoting workload-mitigating strategies before and during episodes of increased workload. One aim of this study is to explore the design of such an operator-based adaptive system, based on monitoring performance and providing support when predicted cascading effects are likely to create downstream workload.

The following three design aspects of such an adaptive system are discussed:

1) a metric for monitoring control actions and their effect on future task demands;
2) the type of support that is provided;
3) the automation's logic for computing conflict resolutions, as needed to provide automated CD&R support.

A. Monitoring Controller Performance

In tactical air traffic control operations, CD&R is characterized by aircraft on straight trajectories. Conflict detection algorithms predict each aircraft's linear trajectory and estimate the separation between two aircraft over time. When two aircraft are closing in, the instance of minimum separation is referred to as the closest point of approach (CPA). For each aircraft pair in an airspace, one can calculate the separation at CPA (d_{CPA}) and the time until the pair will reach this CPA (t_{CPA}). The t_{CPA} is positive for converging aircraft pairs, and negative for diverging pairs.

The U.K.'s National Air Traffic Services (NATS) developed a visual aid called the "separation monitor" [32], a 2-D plot of the d_{CPA} versus the t_{CPA}, with each data point representing an aircraft pair. Fig. 1 shows an example of a traffic scenario (left) and the corresponding separation monitor plot (right). In the latter, converging aircraft pairs are positioned to the right of the *y*-axis, as they have yet to reach their CPA and have a positive t_{CPA}. With time, these aircraft pairs move to the left horizontally as their t_{CPA} reduces.

A critical area can be defined that indicates which aircraft pairs should be closely monitored to prevent a loss of separation (LOS), defined as an instance in which the separation between aircraft is lower than a minimum required distance (5 NM for en-route airspaces). Unless one of the aircraft alters their course, an LOS event will occur when the aircraft pair has $d_{\text{CPA}} < 5$ NM, indicated with a horizontal dashed line in the separation monitor. For aircraft pairs below this line, the time to LOS (t_{LOS}) has been plotted alongside the t_{CPA} (LOS occurs prior to reaching CPA). A second, vertical line indicates a time frame in which action is desirable to prevent a future LOS, which can be thought of as the minimum look-ahead time for resolving conflicts, with 5 min as a typical value [33].

When the controller commands a change in trajectory, e.g., by changing the heading and/or speed of one or multiple aircraft, the pairs are repositioned, indicated with the dotted blue arrows and their new positions by red symbols (this change is not instant because the dynamics of the aircraft will create a transient). Successful conflict resolution actions move aircraft pairs out of the critical area *and* do not move other aircraft pairs into the critical area. If new pairs are moved into the critical area, the action caused secondary conflicts, which decreases airspace stability and creates future additional workload when these conflicts need to be resolved.

Thus, the effectiveness of a controller's action can be assessed in real time by keeping track of the number of aircraft pairs in the critical area before and after the controller's action. An effective action reduces this number. Vice versa, an ineffective action causes an increase. This logic is the basis for the triggering mechanism applied in this article. To account for special types of conflicts for which it is inevitable to induce additional conflicts (and should, therefore, still be counted as effective actions), the system compares the controller's resolution to four other resolution options that are generated by an algorithm. Then, if an action of the controller causes a bigger increase in the number of pairs in the critical area compared to the best of the four resolution options, the system records a self-complicating control action.

B. Design of Support

Automation support for CD&R often involves automatic identification of conflict resolution advisories (RA), with various degrees of controller input and involvement. Our system provides two levels of support to the controller. The low level supports the controller with conflict detection through STCAs. A visual alert is given 130 s before an LOS event, and an additional aural and visual warning is given 1 min before LOS. The high level is triggered after a critical number of self-complicating decisions are recorded, and also supports with conflict resolution through RAs. The controller has 15 s to accept or reject an RA, after which it is implemented automatically.

Fig. 2 shows a schematic of the radar screen with five aircraft, each with a circle indicating the minimum required separation and a label (1) showing its call-sign (ACID), flight level, indicated airspeed, and target waypoint. Only aircraft within the sector (enclosed by the lines connecting the way points) can be controlled. In the upper left corner of the screen (2), a box with experiment information is shown.

When the high level of support is invoked, an aural alert and a bar indicate that automated RAs are provided (3). The interface also shows why the controller's action was considered ineffective by highlighting all aircraft that are involved in any downstream conflict created by the controller. RAs are provided in a pop-up window (4) that shows RAs with an ACID and an expiration time, counting down from 15 to 0 s. Controllers can select an RA, which highlights the aircraft involved and the advised solution (5), and click the buttons to accept or reject the RA.

C. Workload-Mitigating Resolution Advisories

Based on best practices and strategies of air traffic controllers [26], [27], [28], [29], [30], heuristics were used to automatically construct RAs that conform to human reasoning. For example, controllers prefer to vector just one aircraft to resolve a conflict, and solutions are generally considered pairwise first, followed by an assessment of how these solutions fit with the broader traffic situation [33]. High conformance with human reasoning is beneficial for the following two reasons:

1) acceptance of automation increases when the automation uses similar strategies as the operator [2], [3], [34];
2) strategies of expert controllers have already been optimized with respect to workload management, efficiency, and safety, as discussed in the background.

The experiment focused on heading changes to resolve conflicts (i.e., no altitude or speed commands). When considering conflicts pairwise and with vectoring just one aircraft (i.e., dual, complementary resolution actions are not considered), four possible resolutions can be identified: vector aircraft A left or right, or vector aircraft B left or right. Based on these four options, the following best practices were identified based on conflict geometry [33].

1) Overtaking conflicts, with a fast aircraft and a slow aircraft traveling in the same general direction (heading difference of 0 to 45°), and the slow aircraft is ahead of the fast aircraft. A best practice is to vector the fast aircraft around the slow aircraft.
2) Crossing conflicts, with two aircraft on crossing paths (heading difference of 45 to 135°). A best practice is to vector the slower aircraft *behind* the faster aircraft.

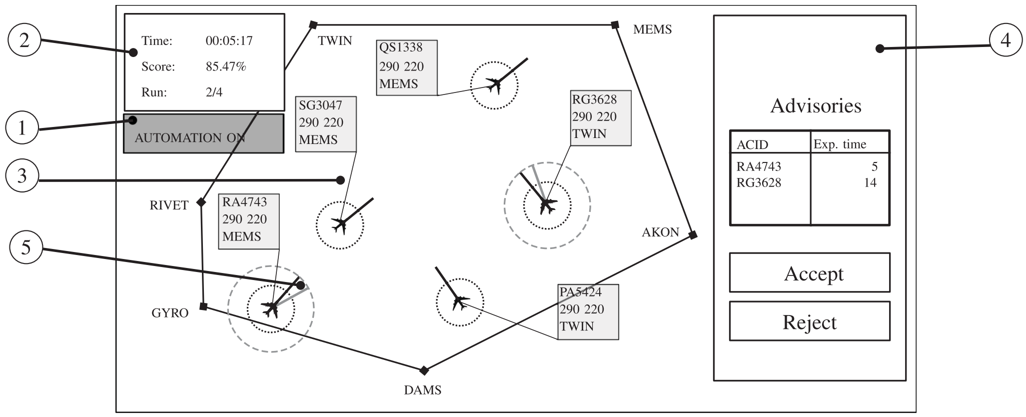

Fig. 2. Schematic of the radar screen with the designed automation support that was used for the experiment. Dimensions are not to scale.

3) Head-on conflicts, with two aircraft traveling in opposite directions (heading difference of 135 to 180°), and no clear best practice. Any aircraft can be vectored left or right.

The algorithm computes the four possible resolutions based on geometry and selects the option that fits these best practices. Just like a human controller, however, the algorithm should reflect that these best practices are only guidelines, i.e., the automation should use these practices as a template but then make its own tradeoff based on contextual factors such as monitoring load and efficiency. For example, for some asymmetrical conflict in which vectoring an aircraft in front of the other requires a notably smaller heading change, it could be more favorable to not follow the best practice. Thus, the best practices have not been hard coded, but have instead been used to develop and tune a cost function. The following four factors were used for the cost function.

1) *Time-to-CPA of the resolution* (t_{CPA}). Because controllers can only be sure that a conflict has been resolved when aircraft start to diverge, the time-to-CPA can be regarded as a measure for the required monitoring after a resolution has been implemented. This has a direct effect on the controller's workload and is thus a factor that should be minimized. For example, minimization favors sending either of two aircraft *behind* another aircraft.
2) *Heading deviation* ($\Delta\Psi$). This should be minimized. This factor is relevant for conflicts with a bias, in which vectoring one aircraft left or right requires a relatively small heading deviation.
3) *Separation-at-CPA* (d_{CPA}). This accounts for crossing conflicts in which one aircraft reaches the conflict point ahead of the other, resulting in a bias for vectoring the aircraft that arrives at the conflict point later.
4) *Indicated airspeed* (IAS). For crossing conflicts, there is a preference for vectoring the slow aircraft. For overtaking conflicts, the controller generally prefers to vector the fast aircraft.

So far these factors consider pairwise resolutions. A fifth factor was added to account for secondary traffic: the change in number of aircraft pairs in the critical area of the separation monitor (ΔN), to make sure that the automation does not select solutions that create secondary conflicts. These five factors have been combined into a cost function used to calculate relative costs for the four resolution options

TABLE I
WEIGHTS FOR THE COST FUNCTION

Conflict type	t_{CPA} (min)	$\Delta\Psi(°)$	d_{CPA} (NM)	IAS(kts)	ΔN
Overtaking	0	0.2	0	-0.02	10
Crossing	0.65	0.02	-0.2	0.01	10
Head-on	0.55	0.1	0	0	10

$$C = w_1 \cdot t_{\text{CPA}} + w_2 \cdot |\Delta\Psi| + w_3 \cdot d_{\text{CPA}} + w_4 \cdot \text{IAS} + w_5 \cdot \Delta N. \tag{1}$$

The weights ($w_1 - w_5$) were tuned such that the algorithm approximates the best practices. For the head-on conflict type, for which there is no best practice, efficiency in minimum heading deviation is the tuning goal. For each conflict type, a specific combination of weights is found, see Table I. The automation selected the resolution option with the lowest cost and presents it as an advisory.

IV. Experiment Design

A. Apparatus

The experiment was conducted in the Air Traffic Management Laboratory (ATMLab), TU Delft. A simulated ATC radar screen was shown on a computer screen; see Fig. 2. In the simulation, participants controlled aircraft using a mouse and keyboard. Instead of radio telephony (RT), aircraft heading commands were given by clicking on an aircraft, setting the aircraft's vector toward the desired heading, and pressing the keyboard ENTER key to send and implement the command. This represents a near-future concept for ATC.

B. Independent Variable

The within-participant independent variable was the triggering threshold for invoking the high level of support. Three different thresholds were tested: conditions AA1, AA2, AA3 represent a low/medium/high threshold condition in which automation support triggers after one, two, or three self-complicating decisions, respectively. Together with a baseline condition with no automation support (MAN), this resulted in a total of four conditions. Additionally, one run was performed with full automation, corresponding to a triggering threshold of zero, and used as a baseline performance condition (AUTO). In AUTO, the automation also vectored aircraft back to their target waypoints as soon as a conflict-free path was available. Because this last condition was fully automated, it was not observed by the participants.

C. Control Variables

1) The revoking strategy (i.e., how and when support returns to its low level), based on duration. A 30-s duration was found effective during the preexperiment testing. After 30 s, any active RAs remained active until it either expired or was accepted or rejected.
2) A 15-s expiration time for RAs.
3) Aircraft altitude was fixed to flight level 290 and could not be changed.
4) Conflict resolutions were fixed to heading clearances.

The latter two control variables increase the likelihood of conflicts and reduces the number of potential solutions compared to current operations. This control was, however, implemented to increase the difficulty of the task (reflecting increased demands that future controllers might experience) and simplify the design of the automation algorithm. In addition, the control was expected to decrease variability in the controller's resolutions, which would ease comparison of the data from multiple controllers and runs.

D. Participants and Instructions

Eighteen participants (2 female and 16 male, average age 27 years) were selected, all students and staff members of TU Delft Aerospace Engineering. All participants had experience with ATC through coursework and were aware of best practices through participating in earlier experiments featuring the same simulation environment [35]. No professional controllers participated because of their limited availability and the fact that the experiment uses a future concept of operation (i.e., the novel interface, no RT but direct manipulation, the adaptive support), which could make it challenging to obtain usable data when experienced controllers are making comparisons to current operations. The experiment consisted of the preexperiment briefing, a training phase, and a measurement phase. During the preexperiment briefing, participants were instructed to execute the following two tasks:

1) maintain a separation of five nautical miles between aircraft at all times;
2) guide all aircraft to their respective exit waypoints in the sector, while minimizing the path deviation and the number of control actions.

There were eight training runs for the participants to learn the simulation environment, the best practices, and the interaction with the adaptive support. The difficulty of the training scenarios

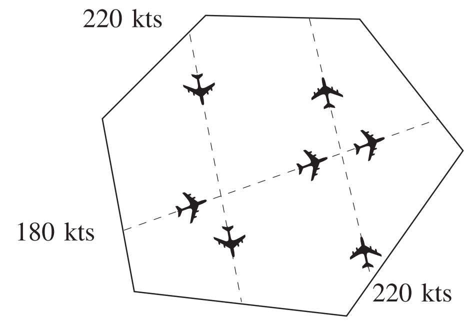

Fig. 3. Sketch of the traffic scenario used for the measurement runs.

increased with each run. Halfway the training, there was one relatively difficult scenario to have participants experience a high task demand. The last three training scenarios were of medium difficulty. Training was followed by one experiment run per condition, with a Latin square distribution to balance any carryover effects across the conditions of interest. Participants were told that the timing of the automation would differ per condition, but they were not aware of the specifics of the triggering threshold and order of the experiment conditions. Breaks were held between the two phases and halfway the measurement runs.

E. Scenario

The traffic scenario was the same for each experiment condition to allow comparison without confound. The airspace was rotated (90, 180, or 270°) between experiment runs such that participants would not recognize the traffic scenario. The sector (see Fig. 3) was about 90 by 90 nautical miles. On average about 11 aircrafts were in the sector at once, which is estimated to be fairly typical for a busy, high-altitude sector such as those in Western Europe [36]. The sector features three intersecting eastbound, northbound, and southbound routes; again, similar to how traffic in a high-altitude sector would be structured. The intersecting traffic flows were perpendicular, resulting in primarily crossing conflict geometry. Nine conflicts were built into the scenario, with the two adjacent intersections requiring participants to consider multiple traffic flows when resolving these conflicts, increasing the likelihood of inducing downstream conflicts.

To test the overall effect of the intervention system and not just the short-term responses, runs were 15-min long. The simulation was run twice the normal speed because of the simplified task (i.e., no RT, etc.). In the first half of the scenario, conflicts emanated at a rate of one conflict per minute, for the second half the rate reduced to one conflict per two minutes. The first two minutes of the scenario were relatively easy to give participants time to adjust.

F. Dependent Measures

Automation triggering is measured as automation operative time, defined as the relative runtime that automation intervention is provided, and the total number of RAs received.

Airspace stability is measured through the domino effect parameter (DEP) [37], a measure for how many new conflicts are created as a result of conflict resolution. This measure is widely used in studies of CD&R performance. New conflicts are defined as the number of stable aircraft pairs with $d_{\mathrm{CPA}} < 5$ NM and $t_{\mathrm{LOS}} < 5$ min [see Fig. 1(b)].

Subjective workload is assessed through instantaneous self assessment (ISA) ratings [38], performed once every minute, and NASA-TLX ratings [39] at the end of each run.

Acceptance of automation support is assessed through controller acceptance rating scale (CARS) ratings [40] after each run, as well as the total number of RAs that were accepted, rejected, or expired.

Efficiency is measured as the average additional track miles flown (compared to a nominal trajectory without CD&R) and the number of control actions. Safety is measured through the number and duration of STCAs and LOS events.

G. Hypotheses

It was hypothesized that conditions with automation support (AA1–AA3) have higher efficiency and safety compared to the baseline condition with no automation support (MAN). Likewise, subjective workload ratings were expected to be lower with automation support compared to MAN.

Additionally, conditions with a low to medium threshold (AA1, AA2) were expected to trigger more interventions, resulting in more pronounced effects on efficiency, safety, and workload compared to higher thresholds (AA3). Automation acceptance was expected to go down with lower thresholds, with AA1 having the lowest acceptance, followed by AA2 and AA3.

V. Results

When the data met conditions for fitting a normal distribution, error bar charts are shown instead of box plots. Repeated-measures ANOVA were performed for data that fit a normal distribution. Although no significant carryover effects were observed, the order in which the conditions were presented to the participants was included as a between-subjects variable in the ANOVAs. Nonparametric Friedman tests were performed when normality assumptions were violated. Bonferroni correction was applied for *post hoc* analysis. Unless statistics are noted, results were not significant.

Data of two participants have been omitted, because their performance was substantially lower than the performance of other participants.

A. Automation Triggering

1) Automation Operative Time: Fig. 4 shows automation operative time. Condition AA1, with the low triggering threshold, has the automation support invoked on average 24% of the time. Higher thresholds (conditions AA2–AA3) reduces the activation of automation support, as would be expected ($F(2, 24) = 31.288, p < 0.001$), significant differences between all conditions.

2) Number of RAs: Table II shows the number of RAs that were accepted, rejected, or expired. Most participants contribute evenly to the total number of RAs, with a few exceptions. As

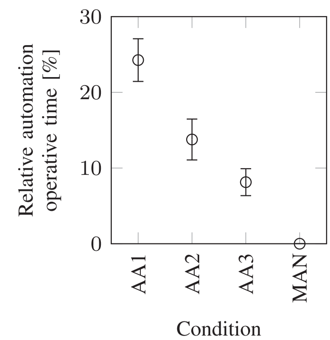

Fig. 4. Automation operative time.

TABLE II
NUMBER OF RAs

Condition	Total	Accepted	Rejected	Expired
AA1	166	127 (76.5%)	7 (4.2%)	32 (19.3%)
AA2	118	79 (66.9%)	6 (5.1%)	23 (19.5%)
AA3	78	60 (76.9%)	5 (6.4%)	13 (16.7%)

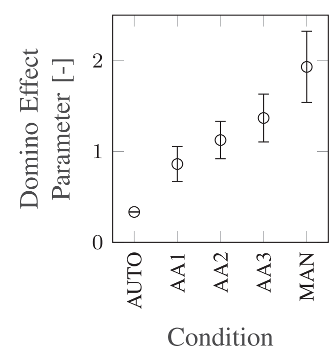

Fig. 5. Airspace stability (higher DEP means lower stability)

expected, in conditions with more support, more RAs were given ($F(2, 24) = 11.105, p < 0.001$), with significant differences between AA1 and AA3.

B. Airspace Stability

Fig. 5 shows the airspace stability, measured as the DEP. A higher DEP reflects a lower stability, with more conflicts created by the controller. To calculate the DEP, the total number of aircraft pairs in the critical area of the separation monitor was divided by the total number of conflicts without CD&R, and reduced by 1 (a bias to set DEP $= 0$ when no additional conflicts are created). Transients of aircraft pairs moving in and out of the critical area when one or more aircraft were making a turn have been filtered out.

A DEP > 0 indicates that additional downstream conflicts were created as a result of the control actions. The data show that the automation (AUTO) solved the traffic scenario while inducing three additional conflicts, with a DEP of 0.33. These downstream conflicts were created beyond the look-ahead time of the automation system. For AA1, on average eight conflicts were self-induced, five more than the AUTO condition, with a

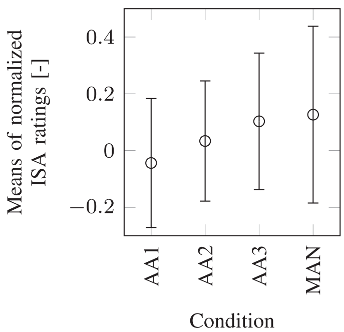

Fig. 6. Mean of ISA ratings.

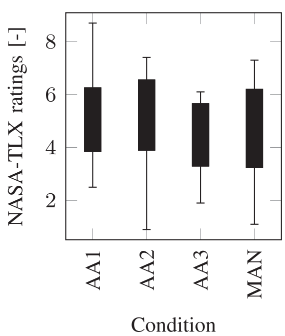

Fig. 7. Total NASA-TLX ratings

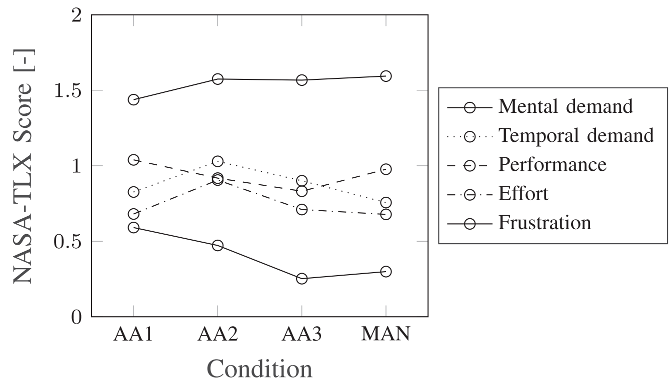

Fig. 8. NASA-TLX scores (rating multiplied with weight).

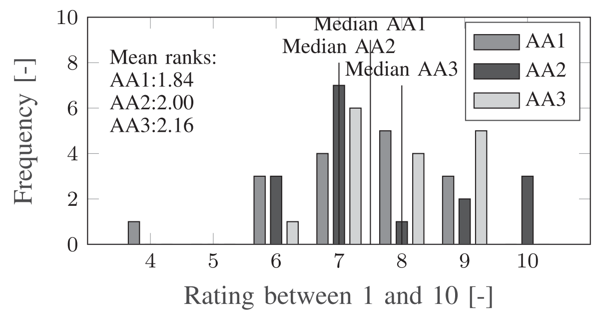

Fig. 9. Controller acceptance ratings (CARS).

mean DEP of 0.86. Considering that AA1 did not allow any self-complicating decisions, had the participants accepted every RA, the condition AA1 would have shown the same number of conflicts as AUTO. Thus, most self-induced conflicts in the condition AA1 can be attributed to the participants rejecting the automation advice. For conditions AA2, AA3, and MAN, more conflicts were self-induced.

This figure confirms that the automation helped to reduce the number of self-induced conflicts and increase airspace stability, as there are notably fewer conflicts with lower threshold automation, and notably lower DEPs ($F(3, 36) = 14.066, p < 0.001$), with significance between AA1 and AA3, AA1 and MAN, and AA2 and MAN.

C. Subjective Workload

1) ISA Ratings: The recorded ISA ratings have been normalized for each participant to obtain a Z-score. The means of the Z-scores over each experiment run are shown in Fig. 6. With low thresholds, ISA ratings are lower, suggesting a small yet nonsignificant decrease in workload.

2) NASA-TLX: The total scores of NASA-TLX ratings are shown in Fig. 7, with no clear or significant pattern. Fig. 8 shows the various components of the NASA-TLX, averaged over all participants. Participants made pairwise comparisons between different components of the NASA-TLX, indicating which were contributing most to their workload, and rated the magnitude of each (on a scale from zero to ten). The pairwise comparisons were used to calculate a weight factor for each component. The product of the weight and the magnitude results in a total score for each component [39]. Because only one participant indicated a contribution of physical demand, this source is not shown.

These data show that the largest contributor to workload is mental demand. AA1 and AA2, with low to medium thresholds, have slightly lower mental demand but more workload originating from effort and frustration. Given that the total NASA-TLX ratings remained fairly constant, while lower thresholds reduces mental demand, this effect is counteracted by increases in effort and frustration.

D. Automation Acceptance

Fig. 9 shows CARS ratings, including the median and mean rank (i.e., how conditions were ranked based on each participant's CARS ratings) for each condition. Conditions with low to medium thresholds show fewer participants rating the automation at an 8 or higher (which would indicate the participants found *"the system satisfactory without improvement"*) and more for 6 or below. The medians are between rating 7 (*"a few improvements are needed"*) and 8 (*"the system is acceptable without improvements"*). The mean ranks decrease with lower thresholds, indicating a slight but nonsignificant decrease in automation acceptance.

Approximately 85% of the RAs were accepted by the participants. Ranking of the data showed that the percentage of accepted RAs decreased slightly but nonsignificantly with lower triggering thresholds.

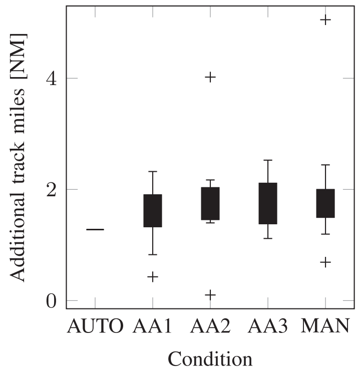

Fig. 10. Additional track miles.

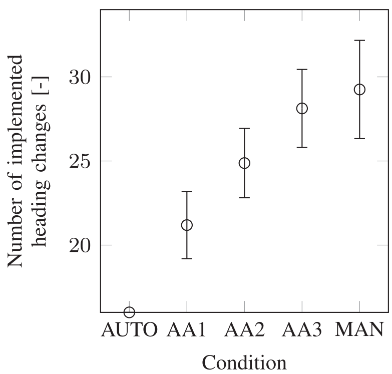

Fig. 11. Implemented heading commands.

TABLE III
MEDIANS AND MEAN RANKS OF THE NUMBER OF STCAS

Condition	Median (interquartile range)	Mean rank
AA1	0 (0 to 2)	1.72
AA2	2 (2 to 2)	2.56
AA3	2 (2 to 2)	2.84
MAN	2 (2 to 2)	2.88

E. *Efficiency and Safety*

1) Additional Track Miles: Although inconsistent, a slight reduction is found with lower thresholds; see Fig. 10.

2) Number of Control Actions: A distinction has been made between the number of *implemented* control actions (i.e., actual heading changes of aircraft) and the *total* number of control actions, including interaction with the pop-up window (accepting/rejecting an RA).

Fig. 11 shows the number of implemented control actions, which decreases with lower triggering thresholds ($F(3, 36) = 5.579, p < 0.01$), with significant differences between AA1 and AA3, and AA1 and MAN.

Fig. 12 shows the total number of control actions. Although the number of implemented control actions decreased with lower triggering thresholds, the total number of actions seems to be constant or even slightly increasing (nonsignificant).

3) STCA and LOS: Table III lists the number of STCAs. The mean ranks show fewer STCAs with lower triggering thresholds ($\chi^2(3) = 15.593, p = 0.001$), with significant differences

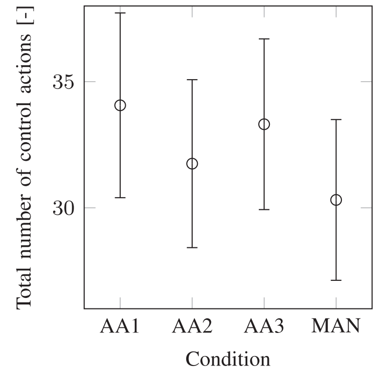

Fig. 12. Total number of control actions.

between AA1 and the three other conditions. Finally, just one LOS event was recorded during an AA1 run, after the participant rejected an RA.

F. *Questionnaire Results*

Participants indicated that they felt the intervention system helped to reduce workload. However, a frequently heard complaint was that the interventions interfered with the participant's own plans, particularly with low and medium triggering thresholds. For example, one participant stated: *"Sometimes I had another plan in my head (looking a step ahead in time)."* Another participant commented: *"It [the automation] proposed a complicated/undesirable situation which felt annoying... [There were] a few stressful situations which were finally solved according to my plan."*

During the experiment, it was observed that participants often put aircraft on conflicting trajectories, commenting that they were aware of downstream conflicts but intended to (re-)vector one of the two aircraft in a different direction in the near future. Thus, they would knowingly create additional conflicts (increasing the DEP measure) to resolve near-term conflicts, with the intent to resolve downstream conflicts later on. The adaptive support did not take into account this intent information, and thus, offered different solutions to the controller. This resulted in participants disagreeing with the automation, which can explain the higher frustration ratings for the lower threshold conditions.

VI. DISCUSSION

Results from this study support the notion of a forward-projecting adaptive system to manage controller workload in CD&R tasks. All conditions with automation support outperformed the manual baseline in the number of control actions that were required to safely manage traffic. Moreover, the number of self-induced conflicts was significantly reduced, meaning the system was successful in preventing future workload spikes in the form of controller-induced conflicts.

The results also confirm that the triggering threshold for an adaptive support system has a significant effect on the human–machine system behavior and performance. Lower thresholds are associated with higher performance and safety. Differences in workload were not significant but showed some trends in ISA ratings and mental demands for the NASA-TLX.

With the tasks and adaptive system tested in this study, the results only marginally support the hypothesis that lower thresholds lead to more pronounced decreases in workload. The results show there is a cost associated with lowering the threshold too far, with lower thresholds increasing the perceived effort and frustration. Based on questionnaire results, higher frustration primarily originates from the intrusiveness of more frequent adaptation. Increases in frustration also correspond to decreases in the automation acceptance ratings. Too frequent interventions are being perceived as too disruptive during critical moments. This echoes earlier findings in studies of adaptive systems. For example, Munro et al. [41] found that an intelligent tutoring system that provides immediate feedback reduces performance compared to a less intrusive system in which students could choose whether and when they wanted to view the feedback.

Vice versa, less-disruptive forms of support can reduce frustration, but are likely to partially negate the positive effects of early intervention (when there is still a chance to prevent cascading conflicts). This suggests that an optimum triggering threshold can be found through a tradeoff of mental demands with effort, frustration, and acceptance. Where this optimum lies depends on the specific weighing of workload, efficiency, and safety, but also other factors, such as sector design, traffic density, and controller proficiency, could play a role.

Frustration with the automation's advice contributed to reduced effectiveness of the support. The support often interfered with the controller's own plans, resulting in disagreements. In these situations—even though the automation's advice was the more-globally optimal solution—the controller preferred his/her own solutions with the intention of later fixing any self-induced conflicts. The RAs were designed to conform to controller's strategies while fostering more strategic resolutions, yet they considered each control action as the final resolution without regard of potential follow-up actions. The controller, however, might think several steps ahead, and can sometimes tolerate self-induced conflicts that can be resolved later. It is also possible that the mental cost associated with switching to a new solution, when the controller already assessed the situation and came up with their own solution, decreases acceptance of the RAs; any follow-up steps or predictions that the controller has made based on their own solution now need to be reevaluated against the new solution. Thus, the "distance" between operator and automation solution can be another factor in the acceptance.

Thus, to improve the adaptive system, it would need to infer and account for the intent of the operator, which illustrates a broader challenge with automation that is involved in decision making in complex environments [2]. With professional controllers likely having more elaborate plans and intent compared to the participants in this study, which would make it even more challenging to infer intent, it is likely that the frustration issue will be aggravated in a more realistic and complex environment. Thus, unless we can better infer the intent in complex work environments, adaptive support at the level of decision-making could be problematic and a more realistic approach is an adaptable system in which the controller decides the level of support. A potentially more fruitful direction for adaptive systems is to support controllers with information integration, with full decision-making authority with the human. For example, visualizing functional constraints through an ecological interface design [42] or showing the separation monitor as an overview of conflicting aircraft pairs could be types of support that are less intrusive to decision-making and their success relies less on inferring intent.

Although findings from this study are specific to the adaptive system designed for this experiment, these issues and the tradeoffs associated with triggering thresholds are likely to generalize to other forms of adaptive systems and application domains. The hope is that similar experiments can be conducted with other adaptive systems, for example, by varying the thresholds for psychophysiological measures such as ECGs. Interestingly, the counterbalancing workload trends of mental demand versus effort and frustration can be a challenge for psychophysiological measures, as the different contributors to workload cannot be distinguished. In addition, analysis of triggering thresholds should also be conducted for revoking automation support [7].

This study has several limitations that should be addressed in future experiments. Several dependent measures (e.g., the ISA ratings) had relatively high variance, which is likely due to divergence in the traffic situations during the experiment: Different decisions made by participants early-on created variation in traffic outcomes, such as complexity of the airspace. Although the long runs provide a higher ecological validity to this study, the high variances prevent making stronger claims backed by statistical evidence.

In addition, although participants had experience with CD&R from participation in earlier experiments, they otherwise had relatively little experience compared to professional controllers. Even though the experiments tasks involved a simplified version of CD&R and significant attention was paid to training the participants with controller best practices, the use of inexperienced participants may have resulted in increased perceived workload ratings and higher frequency of intervention by the adaptive system. Future studies should, therefore, examine the effects of triggering thresholds with more experienced controllers.

Another future research direction includes considering training as an alternate use case for the triggering mechanism. In the training phase of the experiment, the adaptive system was used to verify that participants met a baseline performance threshold. Similarly, it can be used to help novices learn best practices by providing personalized feedback based on their control actions. Further analysis could explore a training tool and investigate whether it can improve learning rate and problem-solving skills compared to existing methods.

Likewise, considering the "distance" between the controller's decision and the automation's suggested conflict resolution as a factor in high frustration, the triggering mechanism could be used to reduce this distance and create a more joint human-automation system. The adaptive automation as a training tool could nudge controllers toward the automation's more global considerations. Likewise, through repeated measures and reinforcement learning, the controller could nudge the automation toward solutions that better account for the intent of the controller.

VII. Conclusion

Experimental results show that the adaptive system prevented self-induced conflicts and increased airspace stability. Lower triggering thresholds resulted in fewer self-induced conflicts, and the operations were more efficient and safe. However, frustration and effort of the controllers increased, and automation acceptance decreased. Fundamental tradeoffs are associated with setting a triggering threshold, and these are important parameters to account for in the design and tuning of adaptive systems. Findings from this study provide, to the authors' knowledge, the first empirical exploration of the trade space associated with this design variable. More research is needed to guide practical implementation of adaptive systems in ATC and other domains facing similar challenges.

References

[1] M. IJtsma, C. Borst, G. A. Mercado-Velasco, M. Mulder, and R. van Paassen, "Adaptive automation based on air traffic controller decision-making," in *Proc. Int. Symp. Aviation Psych.*, 2017, pp. 461–466.

[2] C. Westin, C. Borst, and B. Hilburn, "Strategic conformance: Overcoming acceptance issues of decision aiding automation," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 1, pp. 41–52, Feb. 2016.

[3] B. Hilburn, C. Westin, and C. Borst, "Will controllers accept a machine that thinks like they think? the role of strategic conformance in decision aiding automation," *ATC Quart.*, vol. 22, pp. 115–136, 2014.

[4] W. B. Rouse, "Adaptive allocation of decision making responsibility between supervisor and computer," in *Monitoring Behavior and Supervisory Control*, T. B. Sheridan and G. Johannsen, Eds. New York, NY, USA: Springer, 1976, pp. 295–306.

[5] K. M. Feigh, M. C. Dorneich, and C. C. Hayes, "Toward a characterization of adaptive systems: A framework for researchers and system designers," *Hum. Factors*, vol. 54, no. 6, pp. 1008–1024, 2012.

[6] N. B. Steinhauser, D. Pavlas, and P. A. Hancock, "Design principles for adaptive automation and aiding," *Ergonom. Des.*, vol. 17, no. 2, pp. 6–10, 2009.

[7] C. F. Rusnock and C. D. Geiger, "Simulation-based evaluation of adaptive automation revoking strategies on cognitive workload and situation awareness," *IEEE Trans. Human- Mach. Syst.*, vol. 47, no. 6, pp. 927–938, Dec. 2017.

[8] H. F. R. Arciszewski, T. E. de Greef, and J. H. van Delft, "Adaptive automation in a naval combat management system," *IEEE Trans. Syst., Man, Cybern.*, vol. 39, no. 6, pp. 1188–1199, Nov. 2009.

[9] T. Joo and D. Shin, "Formalizing human-machine interactions for adaptive automation in smart manufacturing," *IEEE Trans. Human- Mach. Syst.*, vol. 49, no. 6, pp. 529–539, Dec. 2019.

[10] F. Alqahtani, S. Katsigiannis, and N. Ramzan, "Using wearable physiological sensors for affect-aware intelligent tutoring systems," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3366–3378, Feb. 2021.

[11] J. Paladines and J. Ramírez, "A systematic literature review of intelligent tutoring systems with dialogue in natural language," *IEEE Access*, vol. 8, pp. 164246–164267, 2020.

[12] C. H. Ting et al., "Real-time adaptive automation system based on identification of operator functional state in simulated process control operations," *IEEE Trans. Syst., Man, Cybern.*, vol. 40, no. 2, pp. 251–262, Mar. 2010.

[13] R. Parasuraman, K. A. Cosenzo, and E. De Visser, "Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload," *Mil. Psychol.*, vol. 21, no. 2, pp. 270–297, 2009.

[14] F. Lillo et al., "Coupling and complexity of interaction of STCA networks," *Proc. Eurocontrol Conf. 8th Innov. Res. Workshop Exhib.*, 2009, pp. 1–8.

[15] R. Göritzlehner, C. Borst, J. Ellerbroek, C. Westin, M. M. van Paassen, and M. Mulder, "Effects of Transparency on the Acceptance of Automated Resolution Advisories," in *Proc. IEEE Int. Conf. Syst., Man, Cybern*. San Diego, CA, USA: IEEE, 2014, pp. 2965–2970.

[16] J. K. Kuchar and L. C. Yang, "A review of conflict detection and resolution modeling methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 1, no. 4, pp. 179–189, Dec. 2000.

[17] K. Reinhardt, M. Poppe, and S. Herr, "Increasing capacity or productivity with controller assistance tools in high complexity airspace," in *Proc. IEEE/AIAA 34th Digit. Avionics Syst. Conf.*, 2015, pp. 1B6-1–1B6-10.

[18] B. Korn, V. Mollwitz, T. Finck, C. Edinger, and S. Tittel, "Validating sectorless ATM in the Hungarian airspace: Results of human in the loop simulations," in *Proc. Integr. Commun., Navigation Surveill. Conf.*, Sep. 2020, pp. 3B1-1–3B1-9.

[19] A. Kuenz, "The 5th dimension in conflict management—XYZT capability," in *Proc. IEEE/AIAA 34th Digit. Avionics Syst. Conf.*, 2015, pp. 1A5-1–1A5-9.

[20] B. Hilburn, P. G. Jorna, E. A. Byrne, and R. Parasuraman, "The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload," in *Human Automation Interaction: Research and Practice*, M. Mouloua and J. M. Koonce, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1997, pp. 84–91.

[21] A. T. Corbett, K. R. Koedinger, and J. R. Anderson, "Intelligent tutoring systems," in *Handbook of Human-Computer Interaction*, 2nd ed., M. Helander, T. K. Landauer, and P. Prabhy, Eds. Amsterdam, The Netherlands: Elsevier Science B.V., 1997, pp. 849–874.

[22] S. F. Scallen, J. A. Duley, and P. A. Hancock, "Pilot performance and preferences for cycles of automation in adaptive function allocation," in *Human Performance in Automated Systems: Current Research and Trends*, M. Mouloua and R. Parasuraman, Eds. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1994, pp. 154–160.

[23] D. B. Kaber and M. R. Endsley, "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task," *Theor. Issues Ergonom. Sci.*, vol. 5, no. 2, pp. 113–153, 2004.

[24] A. Corbett and J. Anderson, "Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2001, pp. 245–252.

[25] M. W. Lewis and J. R. Anderson, "Discrimination of operator schemata in problem solving: Learning from examples," *Cogn. Psychol.*, vol. 17, no. 1, pp. 26–65, 1985.

[26] K. W. Kallus, D. Van Damme, and A. Dittmann, "Integrated task and job analysis of air traffic controllers—Phase 2: Task analysis of en-route controllers," Eurocontrol, Tech. Rep. HUM.ET1.ST01.1000-REP-04, 1999.

[27] J. D'Arcy and P. S. Della Rocca, "Air traffic control specialist decision making and strategic planning—A field survey," Federal Aviation Administration, Atlantic City, NJ, USA, Tech. Rep. DOT/FAA/CT-TN01/05, 2001.

[28] S. Loft, P. Sanderson, A. Neal, and M. Mooij, "Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications," *Hum. Factors*, vol. 49, no. 3, pp. 376–399, 2007.

[29] T. Kontogiannis and S. Malakis, "Strategies in coping with complexity: Development of a behavioural marker system for air traffic controllers," *Saf. Sci.*, vol. 57, pp. 27–34, 2013.

[30] T. L. Seamster, R. E. Redding, J. R. Cannon, J. M. Ryder, and J. A. Purcell, "Cognitive task analysis of expertise in air traffic control," *Int. J. Aviation Psych.*, vol. 3, no. 4, pp. 257–283, 1993.

[31] E. Hollnagel, "Time and time again," *Theor. Issues Ergonom. Sci.*, vol. 3, no. 2, pp. 143–158, 2002.

[32] M. Irfan, M. J. Bull, A. T. Clinch, and S. J. Pember, "Air traffic control," U. S. Patent US9754494B2, 2012.

[33] B. Kirwan and M. Flynn, "Towards a controller-based conflict resolution tool," Eurocontrol, Brussels, Belgium, Tech. Rep. ASA.01.CORA.2.DEL04-A.LIT, 2002.

[34] B. Kirwan and M. Flynn, "Investigating air traffic controller conflict resolution strategies," Eurocontrol, Brussels, Belgium, Tech. Rep. ASA.01.CORA.2.DEL04-B.RS, 2002.

[35] C. Borst, R. M. Visser, M. M. Van Paassen, and M. Mulder, "Exploring short-term training effects of ecological interfaces: A case study in air traffic control," *IEEE Trans. Human-Mach. Syst.*, vol. 49, no. 6, pp. 623–632, Dec. 2019.

[36] *Maastricht Upper Area Control Centre (MUAC) Annual Rep. 2019*, Eurocontrol, Brussels, Belgium, 2020.

[37] J. Krozel, M. Peters, K. D. Bilimoria, C. Lee, and J. S. Mitchell, "System performance characteristics of centralized and decentralized air traffic separation strategies," *ATC Quart.*, vol. 9, no. 4, pp. 311–332, 2001.

[38] T. Hulbert, "A comparison of the 'NASA-TLX' and 'ISA' subjective workload rating techniques," U.K. Civil Aviation Authority Air Traffic Controller Eval. Unit, Bournemouth, U.K., Tech. Rep., 1989.

[39] *NASA Task Load Index User Manual v. 1.0,* Human Performance Research Group, NASA, 1986. [Online]. Available: https://humansystems.arc.nasa.gov/groups/TLX/downloads/TLX.pdf

[40] K. K. Lee, K. Kerns, R. Bone, and M. Nickelson, "Development and validation of the controller acceptance rating scale (CARS)," in *Proc. 4th USA/Europe Air Traffic Manage. RD Seminar*, Santa Fe, NM, USA, 2001, pp. 1–14.

[41] A. Munro, M. Fehling, and D. M. Towne, "Instruction intrusiveness in dynamic simulation training," *J. Comput.-Based Instruct.*, vol. 12, no. 2, pp. 50–53, 1985.

[42] G. A. Mercado Velasco, C. Borst, M. M. van Paassen, and M. Mulder, "Solution space decision support for reducing controller workload in route merging task," *J. Aircr.*, vol. 58, no. 1, pp. 125–137, 2021.

Martijn IJtsma received the M.Sc. degree in aerospace engineering from the Delft University of Technology, Delft, Netherlands, in 2016, and the Ph.D. degree in aerospace engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2019. He is an Assistant Professor with the Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH, USA. His research interests include cognitive systems engineering, human and machine teaming, resilience engineering, and multiagent modeling.

Clark Borst received the M.Sc. (cum laude) and Ph.D. degrees in aerospace engineering from the Delft University of Technology, Delft, The Netherlands, in 2004 and 2009, respectively. He is currently working as an Assistant Professor in Aerospace Human-Machine Systems with the Delft University of Technology. His research interests include developing human-centered aviation automation through the application and empirical evaluation of cognitive systems engineering and ecological interface design principles.

Marinus M. van Paassen (Senior Member, IEEE) received the M.Sc. (cum laude) and Ph.D. degrees from the Delft University of Technology, Delft, The Netherlands, in 1988 and 1994, respectively, both in studies into the neuromuscular system of the pilot's arm. He is an Associate Professor in Aerospace Engineering with the Delft University of Technology, Delft, The Netherlands, working on human-machine interaction and aircraft simulation. His work on human-machine interaction ranges from studies of perceptual processes, haptics and haptic interfaces, and human manual control, to the design of and interaction with complex cognitive systems, applying cognitive systems engineering and ecological interface design for vehicle control.

Prof. Paassen is an Associate Editor for IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.

Max Mulder (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees (cum laude) in aerospace engineering from the Delft University of Technology (TU Delft), Delft, The Netherlands, in 1992 and 1999, respectively, for his work on the cybernetics of tunnel-in-the-sky displays. He is currently a Full Professor and Head of the section Control and Simulation, Faculty of Aerospace Engineering, TU Delft.

Prof. Mulder is an Associate Editor for the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.