

Exploring Short-Term Training Effects of Ecological Interfaces A Case Study in Air Traffic Control

Borst, Clark; Visser, Roeland M.; Van Paassen, Marinus M.; Mulder, Max

DOI

[10.1109/THMS.2019.2919742](https://doi.org/10.1109/THMS.2019.2919742)

Publication date

2019

Document Version

Accepted author manuscript

Published in

IEEE Transactions on Human-Machine Systems

Citation (APA)

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. Article 8735791. <https://doi.org/10.1109/THMS.2019.2919742>

Important note

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

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Exploring Short-Term Training Effects of Ecological Interfaces: a Case Study in Air Traffic Control

Clark Borst, Roeland M. Visser, M.M. Van Paassen, *Senior Member, IEEE* and Max Mulder *Member, IEEE*

Abstract—In many work domains the push toward higher levels of automation raises the concern of diminishing human expertise. Ecological interfaces could help operators in retaining and potentially even in acquiring expertise as they are hypothesized to lead to a deeper understanding of the work domain. This study explores the short-term impact of ecological interfaces on knowledge development and compares the results with an instruction-based training method. To monitor and compare students’ progress, their decision-making strategies, identified from verbal comments recorded in ‘think-aloud’ simulator sessions, are mapped onto the Decision Ladder. This method has been applied to an experiment ($N = 16$) aimed at training novices in conflict detection and resolution (CD&R) within a simplified air traffic control context. Results show that the overall CD&R performance in the final measurement sessions, featuring a transfer manipulation, was not significantly different between the ‘ecological’ and ‘instructional’ groups. In terms of cognitive behavior, however, students in the ecological group exhibited more laborious rule- and knowledge-based behavior that sparked goal-oriented thoughts and corresponding control performances beyond the CD&R task. These findings indicate that ecological interfaces can change how people think and approach a control problem, even after removing the support. It is therefore reasonable to believe that ecological interfaces can play an important role in the early stages of deep knowledge development.

Index Terms—Ecological Interface Design, Training, Human-Machine Interface, Air Traffic Control.

I. INTRODUCTION

MANY work domains are moving toward higher levels of automation to meet more stringent safety, efficiency and productivity demands. As articulated in Bainbridge’s *Ironies of Automation* [1] and in recent work [2], a concern is that the cognitive expertise of human operators will diminish. Ironically, human expertise is critical for handling situations where automation support is unavailable (e.g., due to failures). Ecological Interface Design (EID) could help operators in retaining expertise as ecological displays provide a deeper insight into the physics and causal processes governing their work [3]–[6]. By serving as an “externalized mental model (...) that can support thought experiments and other planning activities” [4, p. 599], could ecological interfaces also contribute to building expertise by shaping the *internal* mental model? Previous longitudinal studies in process control have indeed reported skill and knowledge acquisition of students after being exposed to an ecological interface over a period of six months [7], [8]. Despite the encouraging findings, these studies also reported several limitations.

C. Borst, R.M. Visser, M.M. Van Paassen, and Max Mulder are with the Section Control and Simulation, Delft University of Technology, Delft, The Netherlands. e-mail: c.borst@tudelft.nl

First, students did not receive any initial training, meaning that they had to engage in discovery learning while working with the interface. For some students, however, this incited ‘surface learning’ and led to shallow knowledge as they did not actively reflect on the displayed information [8]. Second, the knowledge acquisition process was monitored by written ‘control recipes’, in which students needed to write down a set of instructions on how they controlled the system. Although this gave insight into how students organized and chunked their knowledge, it cannot be ruled out that a hindsight bias may have confounded these recipes [8].

In this article, a new empirical investigation is described that is aimed at overcoming above-mentioned limitations, whilst focusing on a different application domain: Air Traffic Control (ATC). The scope will be training ATC novices, who are unbiased by previously developed strategies, in a conflict detection and resolution (CD&R) task in the horizontal plane. An ecological interface developed in a previous study, the Solution Space Diagram (SSD) [9], [10], will be used for this purpose. The general approach will be similar to the study conducted by Christoffersen and colleagues [7], [8] in that the control performance and acquired knowledge of two participant groups are tested and compared after a transfer manipulation where the ecological support will be removed. Despite this similarity, there are three important differences.

First, both participant groups received the same initial training by a set of ‘best practice’ instructions in CD&R [11]. This gives participants a head start in proper knowledge development and facilitates a more fair comparison between the two groups. Second, the gained knowledge and control strategies were monitored by mapping verbal comments, recorded during think-aloud simulation sessions, onto Rasmussen’s Decision Ladder (DL) [12]. Thinking aloud while performing a task eliminates a potential hindsight bias. Third, the experiment took place over just two days, thereby aiming to explore short-term effects of EID on training.

The overall goal of the work described in this article is two-fold. First, related to the ATC work domain under investigation, to explore how the SSD contributes to picking up industry ‘best practices’ and facilitates thinking and/or acting beyond those rules. Second, to provide new empirical insights into the merits and versatility of EID on knowledge development of system operators.

II. BACKGROUND

A. Motivation for EID in ATC training

In ATC, trainees are taught to expedite air traffic safely and as efficiently as possible, using a combination of learned

strategies and procedures in response to recognized patterns and conflict geometries [13], [14]. Self discovery of ‘what works’ is typically how they learn the required skills. Trainees are also faced with unexpected disturbances that challenge earlier proven solutions, to discourage a ‘solve-all’ strategy and to encourage knowledge-based problem-solving rather than memorizing ‘tricks’ [14].

Given the description of the ATC training process and its objectives, similarities can be discovered with the tenets of EID. First, similar to how ATC trainees are taught ‘robust’ control strategies instead of fixed procedures, the EID framework was founded on the basic principle of providing support for unanticipated events for which no procedures exist. Ecological interfaces typically do so by portraying the range, or space, of possibilities (e.g., governed by the constraints of the work domain) instead of presenting single optimized solutions that may fall short in situations that violate their specific assumptions [5].

Second, the air traffic controller needs to develop a mental model of the operational situation and how elements of the tactical traffic situation (e.g., aircraft positions, their flight directions, atmospheric conditions, etc.) relate to a higher-level strategic ‘situation’ that contains information about the current and future state of the airspace under control [14]. In EID, Rasmussen’s Abstraction Hierarchy and/or the Abstraction-Decomposition Space serve a similar purpose, by grouping domain-relevant constraints at different levels of abstraction, ranging from lower-level states and whereabouts of objects to their relationships with higher-level functional goals in the operational environment. A goal of EID is to portray this work domain structure on a display to serve as an “externalized mental model” of the system under control [4].

Third, controller expertise is influenced by a large number of perceptual factors. This is not surprising, considering that a controller needs to gain knowledge about the state of the airspace entirely from a Plan View Display (PVD), i.e., the electronic radar display. An ideal ATC training tool should thus support the trainee to become familiar with intricacies of the operational context by actively supporting this action-perception cycle. In this view, ecological interfaces typically aim to transform a cognitive task into a perceptual task [4], potentially enabling users to expedite situation recognition and formulate solutions to problems.

B. Supporting solution strategies for workload mitigation

An important trait of expert controllers is that they manage their own workload by applying solution strategies that minimize the required monitoring time [15]–[19]. Although ATC instructors do not teach particular solution strategies, literature indicates that expert controllers tend to converge to a range of ‘best practices’ with workload-mitigating properties [15]–[19]. Also here, ecological interfaces are expected to support the development of such practices.

To illustrate, consider the Solution Space Diagram (SSD) shown in Figure 1, an ecological interface developed for ATC [9], [10]. In its most succinct form, the SSD portrays velocity obstacles (or, conflict zones) in speed and heading within

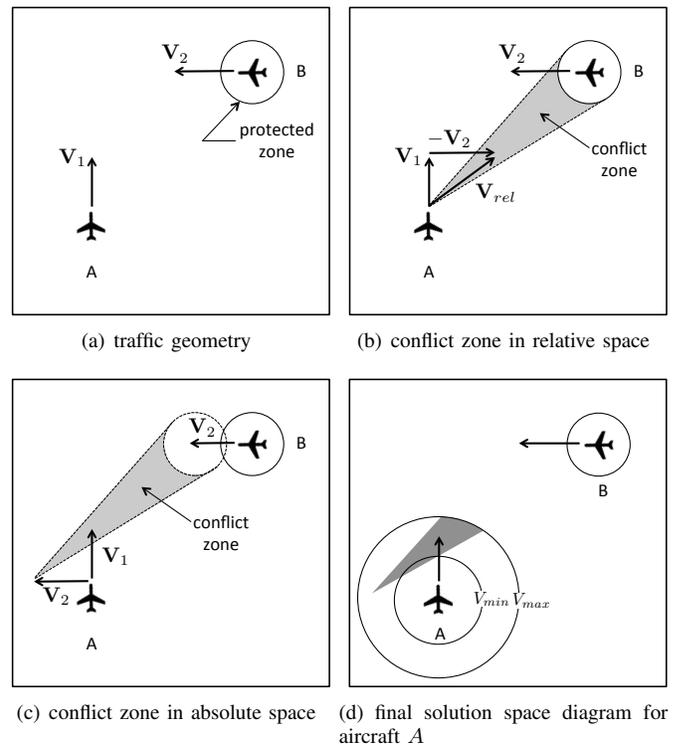


Fig. 1. The Solution Space Diagram (SSD), showing the triangular velocity obstacle (i.e., conflict zone), formed by aircraft B within the speed envelope of the controlled aircraft A, and the absolute speed vectors of both aircraft [adapted from [9], [11]].

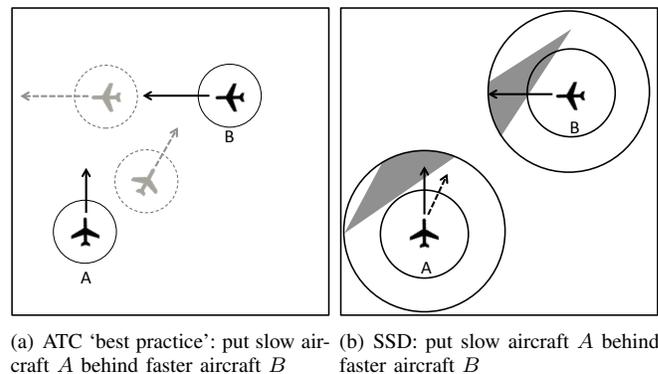


Fig. 2. The SSD explains and re-enforces the ATC ‘best practice’, potentially encouraging the development of control expertise.

the maneuvering envelope of the aircraft under control. The velocity obstacles constrain the maneuvering opportunities of the controlled aircraft in terms of potential loss of separation events. That is, if the velocity vector of the controlled aircraft lies within a triangular conflict zone, a loss of separation will occur in the near future. Vectoring the controlled aircraft outside such a conflict zone resolves the conflict. See [9], [10] for more details on the design.

An example ATC ‘best practice’ to resolve a crossing conflict in the horizontal plane, featuring two aircraft flying at different speeds, is to vector the slow aircraft behind the faster one. This is a typical ‘set-and-forget’ strategy that minimizes monitoring time [19], [20]. Figure 2(b) illustrates why this ‘best practice’ is indeed a robust solution, one that

requires less monitoring. That is, the available solution space on the right hand side of aircraft A 's maneuvering envelope is much richer than on the left hand side. Additionally, placing the speed vector of aircraft A outside the velocity obstacle involves a small heading change to the right, making this a quick solution to resolve the conflict. Hence, the SSD has an explanatory value by making the best practice visually salient. This could stimulate the development of control expertise, because trainees can literally 'see' the 'complete picture' governing a conflict and can thus actively think about and evaluate the best practice.

C. Decision ladder analysis

To support the development of control expertise, one could argue to simply provide ATC trainees with a range of best practices in the form of instructions. We hypothesize, however, the SSD to have certain advantages over instructions alone in the development of control expertise. To illustrate this, an analysis of information-processing steps and resulting knowledge gains for both instructions and the SSD (as illustrated in Figure 2) has been carried out and mapped onto Rasmussen's Decision Ladder (DL) [12].

The DL provides a qualitative model of human decision-making in problem-solving activities and is defined by a sequence of knowledge states (circles) and information-processing actions (boxes), see Figure 3. In general, problem-solving starts at the lower left corner when a worker is confronted with a certain 'problem'. Subsequent information processing would then enable the worker to gain a more detailed understanding of the problem at hand and thus reach higher levels in the DL. After that, several goal-oriented solution options are considered in an iterative cycle of knowledge-based behavior (KBB), followed by a selection of a specific solution and finally leading to the implementation of that solution. The left-hand side of the ladder thus represents problem analysis, whereas the right-hand side constitutes planning and executing solutions. Note that experienced workers rarely follow this sequence in a linear fashion. Based on earlier experiences, they can either skip steps (i.e., knowledge leaps) or make (rule-based) shortcuts between the left- and right-hand sides of the ladder. A shortcut that directly connects the 'activation' and 'execute' boxes of the DL represents skill-based behavior (SBB), which features unconscious automated sensorimotor responses to stimuli. This behavior is commonly associated with highly experienced experts.

Decision-support tools, but also instructions, could support novices to reach higher knowledge states as well as make certain shortcuts salient. For example, in Figure 3(a) it can be seen that best-practice instructions primarily encourage rule-based shortcuts that directly connect learned conflict geometries to task execution. A virtue of rules is that learning them by heart would facilitate swift decision-making, but not necessarily provide support in analyzing the situation. In contrast, the SSD provides more support for the left hand side of the ladder, corresponding to knowledge-based analysis of a traffic situation, see Figure 3(b). It can support novices to reach higher knowledge states in the ladder, toward and into

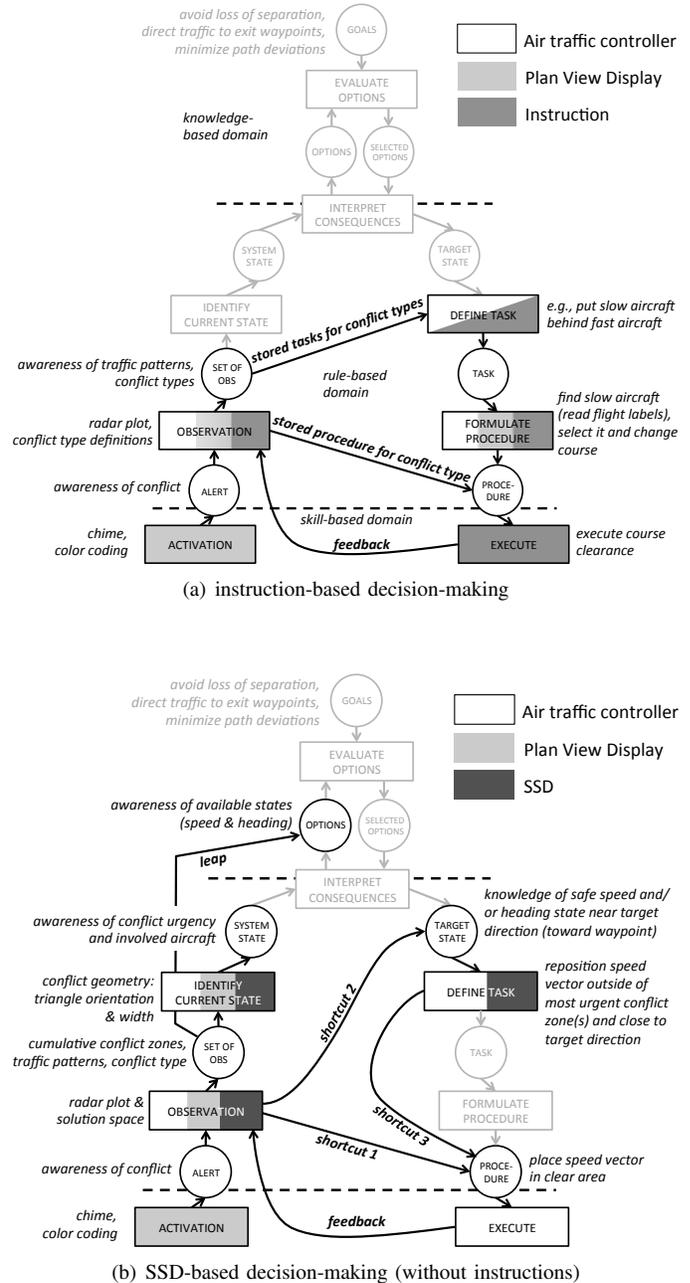


Fig. 3. Decision-making behavior supported by 'best practice' instructions versus the SSD in conjunction with the plan view display.

the region of knowledge-based behavior, whereas instructions would let novices remain more in the rule-based domain.

Decision aides that make rule-based shortcuts salient could result in novices showing expert-like behavior. However, this does not automatically mean novices will acquire the same level of expertise of experienced workers and they may not be able to properly handle situations where the support has been removed. For example, in terms of task execution, the SSD provides several shortcuts from knowledge states toward the definition of tasks and procedures. It would thus depend on the individual user how the SSD will be used in the acquisition of deep knowledge about the traffic situation. That is, the SSD can solely be used as a rule-based tool (shortcuts 1 and 2 in

Figure 3(b)) to resolve a conflict, in which the shortcuts can be formulated as follows: “*Direct the speed vector outside the conflict zone by a heading clearance.*” The risk of using the interface in such a fashion is that it cannot only lead to shallow knowledge and a dependency on the interface, as reported by Christoffersen et al. [8], but also lead to poor control performance (e.g., steer aircraft *A* in Figure 2 in front of aircraft *B*). A person with a more analytical mindset would probably try to reflect on the feedback provided by the SSD (e.g., ask herself a question like “*why is the conflict zone positioned and oriented in this fashion?*”), reach higher regions in the ladder to gain deeper knowledge, and later use that knowledge to fall back to lower-level rule-based shortcuts of better quality.

To facilitate proper decision-making, mitigate potential large variability in SSD usage and encourage a deeper understanding of traffic situations, this study explored a hybrid approach where best practices are taught alongside the SSD. This would also ease measuring changes in achieved (higher-level) knowledge states and help in analyzing to what extent that knowledge would persist after removing the SSD support.

III. EXPERIMENT DESIGN

A. Participants

TU Delft aerospace students were invited to participate voluntarily. After an intake questionnaire (probing their familiarity with ATC goals, displays and practices) and a short skill test (several traffic stills for a conflict detection task), two balanced groups of eight participants (average age of 26 years; standard deviation of 1.9) were formed. All participants were familiar with the existence and looks of ATC radar displays and the overall ATC task (obtained from an introductory course in Avionics), but naive in terms of interpreting the radar display and CD&R best practices.

B. Instructions

All participants were given a mini lecture, in the form of a scripted PowerPoint slideshow, introducing ATC, the PVD and explaining the five best practices to paired aircraft conflicts, see Figure 4 and Table I. These ‘best solutions’ dictated in a specific conflict geometry of an aircraft pair what the best and most efficient action was to solve that conflict. The solutions to the conflicts were distilled from general rules of thumb that were adapted from research about controller strategies (e.g., [18], [20]) and feedback from external experts on ATC training programs. These ‘best solutions’ provide a straightforward and quick fix for a conflicting pair of aircraft.

When practicing in the simulation environment, the task of the participant was to first guarantee safe separation of aircraft at all times by solving or preventing conflicts, and secondly to vector aircraft as efficiently as possible toward their respective exit waypoint. These goals and their priority closely resembled ATC practices.

Finally, participants were instructed to ‘think aloud’ during *all* simulator sessions. Specifically, they were asked to mention the conflict type, the callsign of aircraft involved in the conflict, the aircraft they selected to resolve the conflict

and the type of solution. This allowed us to gain insight in their decision-making strategies and map those onto the DL after the experiment. It also allowed for classification of participants’ decision-making behavior in terms of skill-, rule- and knowledge-based behavior.

C. Independent variables and scenarios

The two independent variables were **training** (two levels), i.e., best-practice instructions with or without the SSD and the **traffic scenarios** (five levels). Training with or without the SSD varied between-participants and only applied to the training phase. After a transfer manipulation, both groups only had access to a baseline PVD to control traffic. Further, training participants in the five conflict types and their corresponding solutions (Figure 4 and Table I) featured aircraft pairs without any other traffic. Traffic scenarios were varied within-participants, and this was realized by changing the order of appearance of the rehearsal exercises. Hence, a mixed design was used. During subsequent training exercises, also scenarios with three aircraft were encountered. Two-aircraft scenarios always had one ‘best solution’, whereas in the three-aircraft scenarios, the third aircraft could either strengthen that ‘best solution’ *or* cause the original ‘best solution’ to create a conflict with the third aircraft. This required the controller to deviate from the learned best practice (see Figure 5).

The traffic scenarios have been developed with the help of two external experts on ATC training (Netherlands), with the specific focus on the horizontal plane and vectoring of aircraft by heading clearances. These restrictions (or simplifications) of the scenarios aimed to prevent confounds caused by the increased number of conflict solution possibilities. Scenarios were classified by the type of conflict (one of the five learning goals), whether the aircraft pair in conflict flew at different speeds, and the number of aircraft surrounding the conflicting pair. The geometrical orientations of the conflicts were rotated or mirrored in order to keep scenarios unrecognizable despite their similar geometries. All traffic scenarios had predefined solutions, such that near-unbiased performance comparisons could be made, and the same feedback could be given afterwards to participants.

Three types of training elements were designed: still conflict scenarios, short dynamic scenarios (90 seconds) and long dynamic scenarios (900 seconds). In the dynamic scenarios, participants could interact with the aircraft and provide heading clearances. The short scenarios were variations of the five learning goals with each a single conflict between two or three aircraft. The longer scenarios were a compilation of at least seven consecutive conflicts, all with different geometries. In these scenarios, multiple foil aircraft were present in the sector to distract the participant, increase complexity and also to force participants to consider multiple solutions when a conflict situation emerged. A new conflict would present itself (turn amber) at least 120 seconds after a previous conflict.

D. Control variables

All traffic was limited to the two-dimensional horizontal plane on flight level 290 and all aircraft were of the same

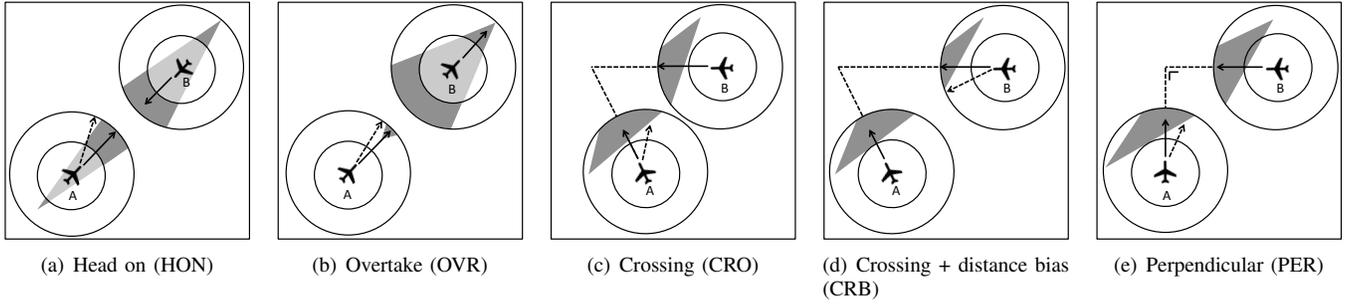


Fig. 4. Conflict types and their visualizations within the SSD. The length of the speed vectors indicate the aircraft speed magnitudes; the dashed speed vectors indicate the best practice solution to the conflict. The dashed line segments indicate the distance toward the crossing point of the aircraft pairs.

TABLE I
CONFLICT TYPES AND THEIR ‘BEST PRACTICES’.

Conflict type	Heading difference [deg]	‘Best practice’ with aircraft speed difference	‘Best practice’ with equal aircraft speeds
Head on (HON)	170 - 180	Faster aircraft evades conflict	Either aircraft, depending on surrounding aircraft
Overtake (OVR)	0 - 10	Overtaking aircraft evades conflict	–
Crossing (CRO)	10 - 170	Slower aircraft evades conflict	Either aircraft, depending on surrounding aircraft
Crossing + bias (CRB)	10 - 170	Aircraft arriving later evades conflict	Aircraft arriving later evades conflict
Perpendicular (PER)	80 - 100	Slower aircraft evades conflict	Either aircraft, depending on surrounding aircraft

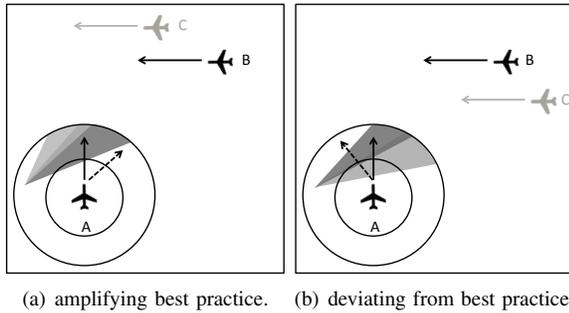


Fig. 5. Example of a three-aircraft scenario where the third aircraft (C) either amplifies, or requires deviation from, the best practice.

type, featuring a speed envelope ranging from 150 kts to 290 kts and a fixed rate-one turning performance. The sector size and shape (squared 50x50 nautical miles area) and waypoint names were constant in all scenarios. To limit the control problem dimensions, conflicts were solved by giving heading clearances to aircraft (i.e., vectoring). No speed and/or altitude changes could be commanded.

E. Dependent Measures

The experiment collected dichotomous performance data during each scenario for three choices made: (1) correct/incorrect conflict recognition, (2) correct/incorrect choice of aircraft and (3) correct/incorrect choice of direction of the solution for the conflict. As the solutions to each conflict problem were pre-defined, simple yes/no answers were noted and cumulative error percentages could be calculated for each participant group. Also, the response times of these three decisions were recorded in seconds (using time-stamped audio and video recordings). In case the recognition of a scenario or the choice for a solution was altered, these would be recorded separately as well. The response time after an initial action was then noted such that in this way the ‘penalty’ time of the

first incorrect choice was included. Other control performance measures included the number of heading clearances (before and after solving the conflict), how often a loss of separation occurred and the total additional flown track miles.

To analyze differences in behavior between the two participant groups, audio and video recordings (a video capture of the computer screen) of the measurement sessions were manually transcribed. To assist the transcription, each information processing step was linked to specific behavioral markers and events observed and measured in the simulation sessions, see Table II. From the DL analysis described in Section II-C and the experiment setup, four most-likely DL traversals were identified and labeled as variants of rule-based behavior (RBB) and knowledge-based behavior (KBB), see Figure 6.

In Figure 6, the first RBB type represents the ‘fastest’ shortcut in which the observation of the traffic scenario immediately leads to an action, irrespective of the correctness of that action. RBB+ involves a more careful identification of the specific conflict type, followed by recalling the procedure involved to act upon the conflict. RBB++ also entails the evaluation of the conflict urgency and the identification of all aircraft involved in (solving) that conflict. Finally, KBB involves ‘thought experiments’ where first (multiple) solutions are evaluated in terms of the higher-order goals (i.e., adhere to target state, avoid new conflicts and minimize path deviations).

F. Apparatus

Aircraft were simulated by linear kinematic equations and described by their position coordinates, velocities and heading angles. Simulations ran on a desktop computer with a 30-inch HD display with a resolution of 2560x1600 pixels and a refresh rate of 60 Hz. Interaction with aircraft was done by direct manipulation using a computer mouse and keyboard. No voice communication was required to command heading changes to aircraft, as this would interfere with the ‘think-aloud’ task of the participants.

TABLE II
IDENTIFICATION OF INFORMATION PROCESSING STEPS.

Step in the DL	Behavioral markers: recognizing these steps during the simulation	Source of the marker
1. Activation	Spots a conflict, attention is drawn towards new conflict	Start of the conflict
2. Observe	Spots all (multiple) aircraft involved	Voice + Video + Cursor
3. Identify	Identifies the type of conflict	Voice
4. Interpret	Considers multiple options as solutions	Voice + Cursor
5. Evaluate	Considers safety of operation and space around aircraft	Voice
6. Define Task	Selects aircraft for the solution	Voice + Video + Mouse click
7. Form Procedure	Selects the direction of the solution	Voice + Video + Mouse click
8. Execute	Executes the solution	Video + Keyboard enter

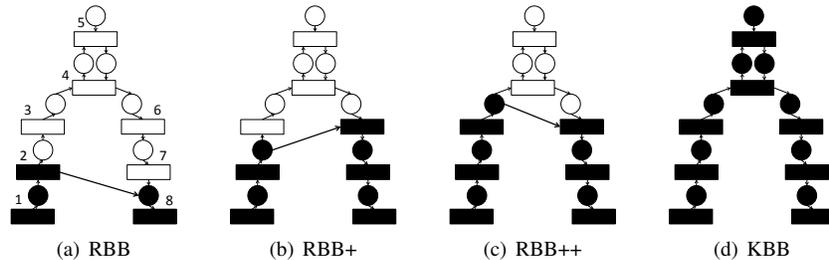


Fig. 6. Decision ladder sequences used to analyze control behavior of participants, ranging from fast rule-based behavior (RBB) toward more laborious and slower knowledge-based behavior (KBB). The numbers in ladder (a) correspond to the information processing steps in Table II.

TABLE III
EXPERIMENT TRAINING PROCEDURE.

Day 1 (morning)		Training Element	Duration
1.	Briefing	Mini lecture (scripted slideshow)	30 min
2.	Training	24 still scenes	25 min
3.	Training	Practice vectoring aircraft	5 min
4.	Training	8 Short dynamic scenes	25 min
5.	Training	2 Long dynamic scenes	30 min
– full-day break –			
Day 2 (afternoon)			
6.	Training	8 Still scenarios (recap, with SSD)	10 min
7.	Training	8 Short dynamic scenes (no SSD)	25 min
8.	Training	2 Long dynamic scenes (no SSD)	30 min
9.	Measurement	5 Short dynamic scenes (no SSD)	15 min
10.	Measurement	1 Long dynamic scenes (no SSD)	15 min
11.	Debriefing	Retrospective questionnaire	20 min

G. Procedure and data analysis

For each participant, the experiment was divided in two half-day sessions (including breaks), with exactly one day in between and balanced to different times during the day (morning or afternoon), see Table III. On the first half day, participants received the 30-minute mini lecture. For the ecological group, an additional slide was presented that explained the SSD and the five best practices were shown in conjunction with the SSD.

The exercises following the briefing first featured 24 still scenarios with just two aircraft that were either in conflict or not. For the SSD group, the SSDs for both aircraft were shown. To prepare for the dynamic scenes in which participants could interact with aircraft and resolve conflicts, a short 5-minute session was dedicated to vectoring an aircraft using the mouse cursor device and the ENTER key. Day 1 was concluded with eight short dynamic scenes, each lasting 90 seconds – first four scenes contained two aircraft and last four scenes contained three aircraft – and two long scenarios. Here, the

short dynamic scenes with three aircraft all reinforced the learned best practice. Note, the ecological group always had access to the SSD in addition to the baseline PVD during all still and dynamic exercises, whereas the instructional group only had access to a baseline PVD.

Day 2 started with a short recap exercise featuring still scenarios with just two aircraft. After that, the transfer manipulation was done for the SSD group in which the remaining exercises (8 short dynamic scenes and 2 long dynamic scenes, same as in Day 1) featured a baseline PVD. The final measurement session consisted of 5 short dynamic scenes, all with three aircraft and five conflict types that either reinforced the best practice or required deviation (see Results section).

In analyzing the results, it was decided to omit the data analysis from the long scenarios, because the large variability in the evolution of traffic situations generally makes the results very difficult to compare between participants and groups. A total of two hours of video and audio recordings was analyzed. First, all video and audio material were anonymized by removing any reference to a particular participant group, allowing for a double blind analysis. After that, all material was transcribed. The transcription was carried out by two evaluators (both familiar with the DL) to ensure unbiased results as much as possible. Given the relatively low sample size for each experimental condition, conservative non-parametric tests were used to compare the control performance between the two participant groups. Kruskal-Wallis and Friedman tests were applied to analyze between- and within-group effects, respectively.

H. Hypotheses

First, it was hypothesized that training with the SSD would incite more higher-level cognitive behavior (i.e., RBB++ and KBB), because it encourages participants to attend to the

complete traffic situation instead of just the conflict pair and the specific rules. In other words, training without the SSD would encourage participants to only ‘find the right rule’ to solve a conflict, whereas training with the SSD was expected to encourage participants in gaining insight into the situation, evaluate the learned rules and pick the best one. Second, the SSD group was expected to make less mistakes in conflict type recognition, choosing the correct aircraft to solve the conflict and implementing the correct solution, albeit at the cost of higher response times. Third, in ‘novel’ scenarios that required deviations from the best practice due to the presence of a third aircraft, the SSD group was expected to develop ‘new rules’ and better handle those situations than the ‘instructional’ group.

IV. RESULTS

In the short scenario sessions, no loss of separation events occurred as all participants managed to keep aircraft separated more than 6 nautical miles. Two participants were removed from the analysis, because they caused significant outliers and showed deviating behavior. Fortunately, they belonged each to a different group, yielding two balanced groups of each seven participants.

A. Conflict detection and resolution

The control performance results, shown in Figure 7, reveal trends in support of the second and third hypotheses, with overall larger response times for the SSD group and observed improvements for this group in the ‘novel’ traffic scenarios (i.e., HON and PER). However, Kruskal-Wallis tests did not find any significant difference between the two participant groups on all aspects of the CD&R control performance.

Friedman tests did find significant differences for the within-group manipulations, i.e., the conflict types. A significant effect of conflict type on the recognition response time was found ($\chi^2(4) = 19.476, p < 0.01$). Pair-wise comparisons (with Bonferoni correction) revealed that the PER scenario had significant longer response times than the HON scenario, despite that both were ‘novel’ scenarios. Also a significant increase in aircraft choice response time was found for the SSD group in conflict type recognition ($\chi^2(4) = 15.928, p = 0.03$). Pair-wise comparisons revealed that CRB was significantly different from OVR and PER. The solution response time showed similar results, with a significant effect for conflict type ($\chi^2(4) = 29.789, p < 0.01$). Here, pair-wise comparisons showed a significant difference between CRO and CRB, OVR and CRB, and PER and CRB.

B. Control efficiency

Figure 8 shows the total number of heading clearances and the total additional track miles flown by aircraft. The heading clearances were split by clearances before (i.e., commanding an evasive maneuver) and after (i.e., commanding aircraft back toward their target waypoint) the conflict. The additional track miles were determined by the flown distances relative to the shortest paths toward the exit points.

In general, the two participant groups did not perform significantly different in terms of number of heading clearances. However, there was a significant group effect on the total additional track miles across the majority of conflict types (HON: $H(1) = 3.922, p = 0.048$; OVR: $H(1) = 9.016, p < 0.01$; CRO: $H(1) = 4.445, p = 0.035$; PER: $H(1) = 4.446, p = 0.035$). There was also a significant difference in track miles between conflict types ($\chi^2(4) = 48.514, p < 0.01$), which is not surprising given the different traffic geometries. The reduction in additional track miles (see Figure 8(b)) for the SSD group indicates that the heading clearances were of better quality by adhering more closely to the aircraft shortest routes toward the exit waypoints. Interestingly, the CRB conflict type was the only one that resulted in similar additional track miles for the two groups. In this scenario, several participants in the SSD group had a tendency to put aircraft on their designated course too early after solving the conflict, occasionally requiring corrective actions to stay clear of the conflict that was initially solved. The instructional group showed less corrective heading adjustments, especially after the conflict was solved. They either waited longer before clearing aircraft to their exit waypoints or not do this at all. This resulted in slightly less heading clearances after the conflict, but more additional track miles.

C. Decision-making behavior

In Figure 9, two transcripts from two different participants are shown for the CRO scenario along with their mapped behavior. As can be seen in this figure, it was common to find more than one behavioral type within one trial per participant. For example, Figure 9(b) indicates that the participant first engaged in KBB by evaluating a possible solution, but later, when executing a second control action, a rule-based shortcut was made representing RBB from Figure 6(a). Instead of counting this as both KBB and RBB, we decided to only count the first identified behavior (i.e., KBB) as we believe this to be most telling and meaningful of how a participant started to approach a (new) scenario. Additionally, it was also common to find iterations within one behavioral type, e.g., several attempts to detect the correct conflict type (RBB+), which were counted separately.

The resulting distribution in decision-making behavior, ranging from fast RBB toward slower KBB (with and without observed iterations), is provided in Figure 10. From this figure it can be observed that, overall, the SSD group reached higher in the DL than the instructional group, as hypothesized. Participants who trained with the SSD evaluated potential solutions against secondary ATC goals (e.g., adhering to target waypoints and avoiding new conflicts) rather than just on solving the conflict. This can also be observed from the transcripts provided in Figure 9, where Figure 9(b) is from a participant in the SSD group. The reduced additional track miles for the SSD group also supports this observation.

Participants in the SSD group consistently showed more RBB++ and KBB counts than participants in the instructional group, irrespective of conflict type. This was not only the case in the ‘easier’ scenarios (e.g., see the CRO transcripts

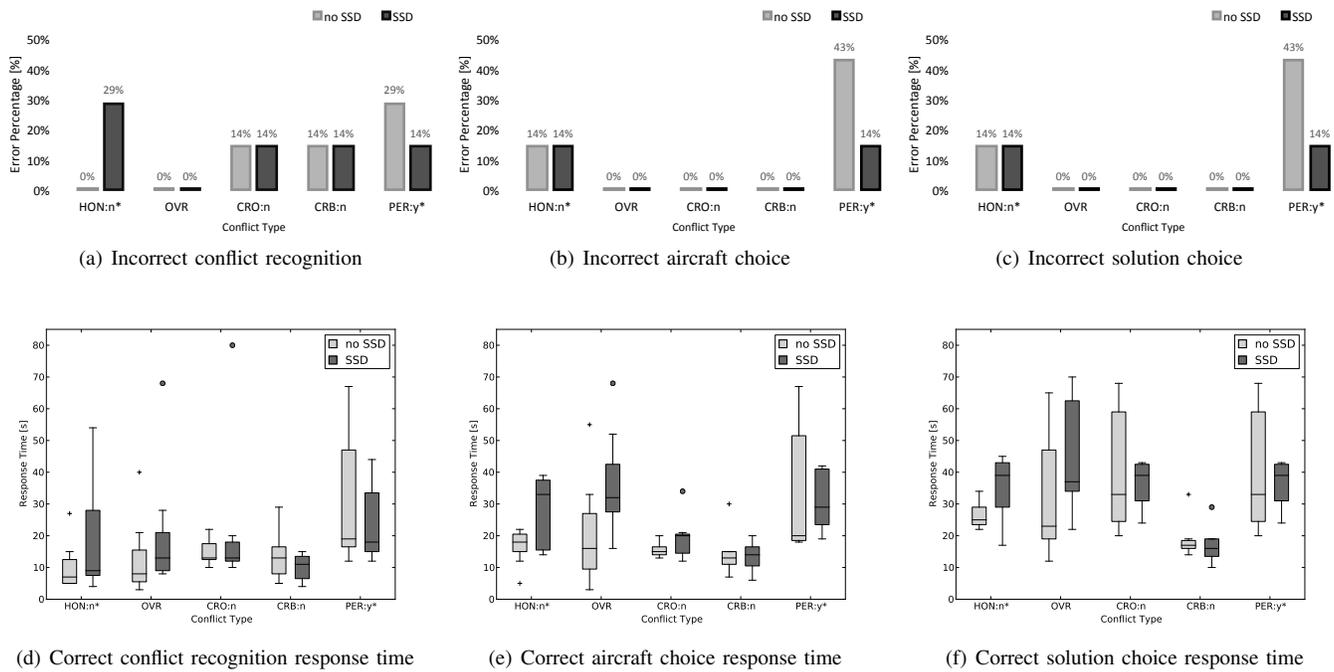


Fig. 7. Cumulative error percentages and response times of conflict type recognition, aircraft choice and solution response. The * symbol indicates deviation from the best practice and ‘y’ or ‘n’ designates the presence of a speed difference ‘yes’ or ‘no.’

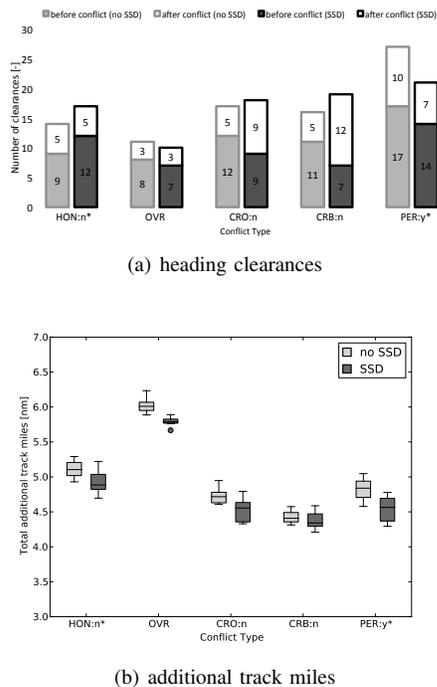


Fig. 8. Total number of heading clearances and additional track miles. The * symbol indicates deviation from the best practice and ‘y’ or ‘n’ designates the presence of a speed difference ‘yes’ or ‘no.’

in Figure 9), but also in the novel PER scenario as shown in Figure 10(b). In this scenario, the SSD group did make less mistakes compared to the instructional group. Note that none of the participants showed the lowest level of RBB behavior. This type of behavior is commonly reserved to highly experienced controllers, and the short duration of our

experiment apparently did not allow students to reach that level of expertise.

V. DISCUSSION

The gist of quantitative and qualitative results suggests that the SSD changed how participants thought and approached their control problem, even after removing the ecological support. Although no significant differences were observed on the primary task performance (i.e., CD&R), the control efficiency and decision-making results indicate that the SSD group was also more attentive to secondary ATC goals (i.e., adhering to target waypoints and minimizing path deviations) and better handle the ‘novel’ PER scenario. Similar functional and goal-oriented behavior and knowledge organization was found by Christoffersen *et al.* [8].

From these results, however, the long-term impact and benefits for employing ecological interfaces in (ATC) training are difficult to predict. For example, specific to ATC experts is that they try to mitigate their own cognitive workload, but the SSD group revealed longer response times and more KBB. Although these results may be explained by the increased attentiveness to secondary ATC goals, it cannot be ruled out that also some confusion, caused by potential “compatibility issues” [21], [22] arisen after the transfer manipulation, contributed to the increased cognitive load. Only a longitudinal study would be able to shed light on how far this effect would persist after a prolonged period of time.

The conflict type effects for the SSD group are also interesting, suggesting the SSD is context sensitive (regarding learning). In Figures 4(a) and (b), the SSD patterns for HON and OVR conflicts make the best practice solution indeed less salient than for the crossing conflict types, given the orientation of the conflict zones. However, these results are largely

Time	Event / DL step	Verbal comment	Mapped behavior
0 s	Start of scenario		
2 s	Alert (chime)	"Ok, we have a conflict"	
13 s	Observe + identify Conflict recognized	"It's a crossing conflict with equal speeds"	
14 s	Define task + form proc. Aircraft selected	"Going to send aircraft SM behind aircraft RA"	
22 s	Execute Solution execution		

(a) Best matching behavior type: RBB+

Time	Event / DL step	Verbal comment	Mapped behavior
0 s	Start of scenario		
2 s	Alert (chime)		
5 s	Observe	"Think RA and SG are in crossing conflict, not sure."	
16 s	Identify Conflict recognized	"No, SM is causing problems. I see no speed difference."	
20 s	Define task Aircraft selected	"Going to send aircraft SM behind both RA and SG"	
22 s	Interpret Evaluate	"SM then stays clear of both aircraft and can be put back later."	
28 s	Form proc. + Execute Solution execution		
32 s	Observe + execute 2 nd command	"Reduce heading change a bit for performance"	
38 s	Observe + execute Clear to exit waypoint	"RA is not aligned with waypoint"	
41 s	Observe	"Yes, all is working out!"	

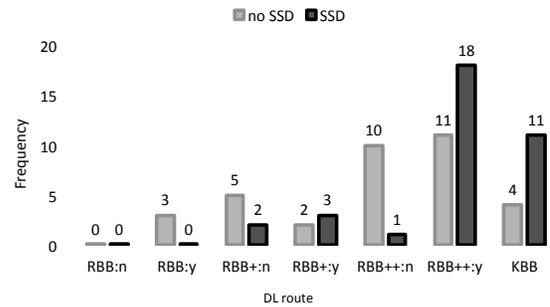
(b) Best matching behavior type: KBB

Fig. 9. Example transcripts and their matching behavior types for the crossing (CRO) scenario.

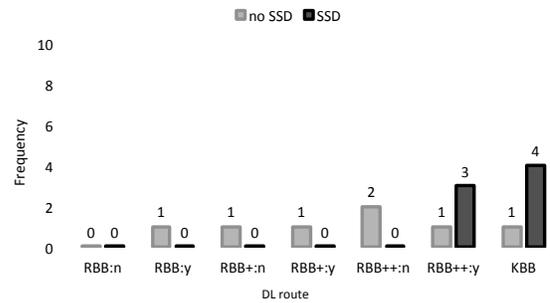
consistent with an earlier study in ATC conflict detection, which showed that conflicts with "large convergence angles" and "short conflict times" (i.e., HON) and "small angles" and "long times" (i.e., OVR) are more difficult to detect than conflicts with "small angles" and "short conflict times" (i.e., CRO) [23]. In that sense, the SSD did not trivialize an intrinsically difficult problem, which is a common experience with ecological displays [5].

Our study also had limitations that need to be mentioned and addressed in future studies. First, transcribing the audio and video material was a laborious process. Although this process was carried out by two evaluators in a double blind fashion, the results depended on how well participants articulated their decision-making behavior and how this was interpreted. In addition, the verbal comments did not always follow a chronological order relative to the sequence of steps in the DL. No reliability test was undertaken by a separate analyst, which could be identified as a limitation in our exploratory study. For future work, developing a set of standardized verbal protocols, matching the steps in the decision ladder, could make this process more streamlined.

Second, we did not take into account specific instructional design methods in teaching students to comprehend and take full advantage of the SSD. After a brief explanation of the SSD and the best practices, students needed to rely on some



(a) Overall decision-making behavior



(b) Decision-making behavior in PER:y*

Fig. 10. Histogram of the observed decision-making behavior, where 'y' or 'n' indicates the presence of iterations within the behavior type.

form of discovery learning. It cannot be ruled out that personal differences in learning style affected our results. Other studies have indeed shown that people who are 'holists' tend to pick up information from an ecological interface more easily than 'serialists' [24]. For future studies, it is recommended to include an instructional design method (e.g., 'visual scaffolding' [25]) that complements ecological interfaces to better guide the information pickup and learning process.

To conclude, the findings in our study have shed light on broader implications for EID in terms of training requirements and its potential role in addressing expertise degradation. As mentioned by Christoffersen *et al.* [8], an ecological interface only contributes to *proper* knowledge acquisition when operators actively reflect on its visual feedback. Our experiment was specifically geared toward training, thereby creating a learning environment that encouraged evaluation and explorative thought experiments next to the industry "gold standards." This appeared to have effect, implying that ecological displays would *always* require training before users can take advantage of the visual feedback in a way that can build and/or retain expertise.

VI. CONCLUSION

We investigated the short-term effects of training a group of novices (in air traffic conflict detection and resolution) with an ecological interface and compared the control performance and decision-making behavior with a group that only received instructions. An experiment was conducted wherein two groups underwent a two-day training program featuring a transfer manipulation in the final measurement scenarios on the second day. Results show that the primary task performance between the 'ecological' and the 'instructional'

group was not significantly different. Interestingly, students in the ecological group exhibited more laborious rule- and knowledge-based behavior that sparked goal-oriented thoughts and corresponding control performances beyond the primary task. These findings indicate that ecological interfaces can change how people think and approach a control problem, even after removing the support. It is therefore reasonable to believe that ecological interfaces can play an important role in the development of deeper knowledge.

REFERENCES

- [1] L. Bainbridge, "Ironies of Automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.
- [2] B. Strauch, "Ironies of Automation: Still Unresolved After All These Years," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 5, pp. 419–433, 2018.
- [3] K. J. Vicente and J. Rasmussen, "The Ecology of Human-Machine Systems II: Mediating Direct-Perception in Complex Work Domains," *Ecological Psychol.*, vol. 2, no. 3, pp. 207–249, 1990.
- [4] —, "Ecological Interface Design: Theoretical Foundations," *IEEE Trans. on Systems, Man, and Cyb.*, vol. 22, no. 4, pp. 589–606, 1992.
- [5] C. Borst, J. M. Flach, and J. Ellerbroek, "Beyond Ecological Interface Design: Lessons From Concerns and Misconceptions," *IEEE Trans. on Human-Mach. Syst.*, vol. 45, no. 2, pp. 164–175, 2015.
- [6] R. C. McIlroy and N. A. Stanton, "Ecological Interface Design Two Decades On: Whatever Happened to the SRK Taxonomy?" *IEEE Trans. on Human-Mach. Syst.*, vol. 45, no. 2, pp. 145–163, 2015.
- [7] K. Christoffersen, C. N. Hunter, and K. J. Vicente, "A Longitudinal Study of the Effects of Ecological Interface Design on Skill Acquisition," *Human Factors*, vol. 38, no. 3, pp. 523–541, sep 1996.
- [8] —, "A Longitudinal Study of the Effects of Ecological Interface Design on Deep Knowledge," *Int. J. of Human-Computer Studies*, vol. 48, no. 6, pp. 729–762, 1998.
- [9] C. Borst, V. A. Bijsterbosch, M. M. van Paassen, and M. Mulder, "Ecological interface design: supporting fault diagnosis of automated advice in a supervisory air traffic control task," *Cognition, Technology & Work*, vol. 19, no. 4, pp. 545–560, nov 2017.
- [10] S. B. J. van Dam, M. Mulder, and M. M. van Paassen, "Ecological interface design of a tactical airborne separation assistance tool," *IEEE Trans. on Systems, Man and Cyb., Part A*, vol. 38, no. 6, pp. 1221–1233, 2008.
- [11] C. Borst, R. M. Visser, M. M. van Paassen, and M. Mulder, "Ecological Approach to Train Air Traffic Control Novices in Conflict Detection and Resolution," in *Proc. Sixth SESAR Innovation Days*, no. November, 2016, pp. 1–9.
- [12] J. Rasmussen, *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering*. North-Holland, 1986.
- [13] Federal Aviation Administration, "Review and Evaluation of Air Traffic Controller Training at the FAA Academy," U.S. Department of Transportation, Tech. Rep., 2013.
- [14] M. Schuver-van Blanken, "Clarifying Cognitive Complexity and Controller Strategies in Disturbed Inbound Peak ATC Operations," in *Advances in Aviation Psychology*, M. A. Vidulich, P. S. Tsang, and J. M. Flach, Eds. Ashgate, 2014, ch. 6, pp. 85–102.
- [15] J. D'Arcy and P. Della Rocco, "Air Traffic Control Specialist Decision Making and Strategic Planning-A Field Survey," FAA, DOT/FAA/CT-TN01/05, Washington D.C., Tech. Rep., 2001.
- [16] S. Loft, S. Bolland, M. S. Humphreys, and A. Neal, "A Theory and Model of Conflict Detection in Air Traffic Control: Incorporating Environmental Constraints," *J. Exp. Psychol.: Applied*, vol. 15, no. 2, pp. 106–124, jun 2009.
- [17] E. M. Rantanen and A. Nunes, "Hierarchical Conflict Detection in Air Traffic Control Hierarchical Conflict Detection in Air Traffic Control," *The Int. J. of Aviation Psychology*, vol. 15, no. 4, pp. 339–362, 2005.
- [18] E. M. Rantanen and C. D. Wickens, "Conflict Resolution Maneuvers in Air Traffic Control: Investigation of Operational Data," *The Int. J. of Aviation Psychology*, vol. 22, no. 3, pp. 266–281, jul 2012.
- [19] S. Fothergill and A. Neal, "Conflict-Resolution Heuristics for En Route Air Traffic Management," in *Proc. of the Human Factors and Ergonom. Soc. 57th Ann. Meeting*, vol. 57, no. 1. SAGE Publications, 2013, pp. 71–75.
- [20] B. Kirwan and M. Flynn, "Investigating Air Traffic Controller Conflict Resolution Strategies," EUROCONTROL, Brussels, Belgium, Rep. ASA.01.CORA.2.DEL04-B.RS, March 2002.
- [21] N. Lau, G. A. Jamieson, G. Skraaning, and C. M. Burns, "Ecological Interface Design in the Nuclear Domain : An Empirical Evaluation of Ecological Displays for the Secondary Subsystems of a Boiling Water Reactor Plant Simulator," *IEEE Trans. Nuclear Science*, vol. 55, no. 6, pp. 3597–3610, 2008.
- [22] Y. Li, X. Wang, and C. M. Burns, "Improved Monitoring Performance of Financial Trading Algorithms Using a Graphical Display," in *Proc. Human Factors and Ergonom. Soc. - 2018 Annual Meeting*, 2018, pp. 187–191.
- [23] R. W. Remington, J. C. Johnston, E. Ruthruff, M. Gold, and M. Romera, "Visual Search in Complex Displays: Factors Affecting Conflict Detection by Air Traffic Controllers," *Human Factors*, vol. 42, no. 3, pp. 349–366, sep 2000.
- [24] G. L. Torenvliet, G. A. Jamieson, and K. J. Vicente, "Making the Most of Ecological Interface Design: The Role of Individual Differences," *Applied Ergonomics*, vol. 31, no. 4, pp. 395–408, 2000.
- [25] J. J. G. van Merriënboer, P. A. Kirschner, and L. Kester, "Taking the load off a learner's mind: Instructional design for complex learning," *Educational Psychologist*, vol. 38, no. 1, pp. 5–13, 2003.



Clark Borst received the MSc (2004, cum laude) and PhD degrees (2009) in aerospace engineering from the Delft University of Technology, The Netherlands, where he is currently working as an Assistant Professor in aerospace human-machine systems. His work and research interests lie in developing human-centered aviation automation through the application and empirical evaluation of Cognitive Systems Engineering and Ecological Interface Design principles.



Roeland M. Visser received the MSc degree (2016) in aerospace engineering from the Delft University of Technology, The Netherlands, for his work in ecological interface design and training for Air Traffic Control. He is currently working as Airport Consultant for Netherlands Airport Consultants (NACO) on strategic and tactical studies which include operational analyses, air traffic forecasting, flight schedule predictions, organisation structuring and operational readiness for airports around the world.



Marinus (René) M. van Paassen (M'08–SM'15) received the M.Sc. and Ph.D. degrees in aerospace engineering from the Delft University of Technology (TU Delft), Delft, The Netherlands, in 1988 and 1994, respectively, for his studies on the role of the neuromuscular system of the pilot's arm in manual control. He is currently an Associate Professor with the section Control and Simulation, Aerospace Engineering, TU Delft, working on human-machine interaction and aircraft simulation. His work on human-machine interaction ranges from studies of perceptual processes and manual control to complex cognitive systems.

Dr. van Paassen is an Associate Editor for the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.



Max Mulder (M'14) 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 the Head of the section Control and Simulation, Faculty of Aerospace Engineering, TU Delft. His research interests include cybernetics and its use in modeling human perception and performance, and cognitive systems engineering and its application in the design of "ecological" human-machine interfaces.

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