Situation Awareness Prompts
Bridging the Gap between Supervisory and Manual Air Traffic Control

Kim, Munyung; Borst, Clark; Mulder, Max

DOI
10.1016/j.ifacol.2022.10.224

Publication date
2022

Document Version
Final published version

Published in
IFAC-PapersOnline

Citation (APA)
Kim, M., Borst, C., & Mulder, M. (2022). Situation Awareness Prompts: Bridging the Gap between Supervisory and Manual Air Traffic Control. *IFAC-PapersOnline*, *55*(29), 13-18. https://doi.org/10.1016/j.ifacol.2022.10.224

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-29 (2022) 13–18

Situation Awareness Prompts: Bridging the Gap between Supervisory and Manual Air Traffic Control

Munyung Kim, Clark Borst [1] and Max Mulder
Control and Simulation, Faculty of Aerospace Engineering, TU Delft, 2629 HS, Delft, The Netherlands

Abstract: To meet increasing safety and performance demands in air traffic control (ATC), more advanced automated systems will be introduced to assist human air traffic controllers. Some even foresee complete automation, with the human as a supervisor only to step-in when automation fails. Literature and empirical evidence suggest that supervising highly-automated systems can cause severe vigilance and complacency problems, out-of-the-loop situation awareness and transient workload peaks. These impair the ability for humans to successfully take over control. In this study, situation awareness prompts were used as a way to keep controllers cognitively engaged during their supervision of a fully automated ATC system. Results from an exploratory human-in-the-loop experiment, in which eight participants were instructed to monitor a fully automated ATC system in a simplified ATC context, show a significant decrease in workload peaks following an automation failure after being exposed to high-level SA questions. Although the selected method did not necessarily yield improved safety and manual control efficiency, results suggest that using situation awareness feedback in line with controllers' attention could be an avenue worth exploring further as a training tool.

Keywords: Decision making and cognitive processes, Human centred automation, Shared control, cooperation and degree of automation

1. INTRODUCTION

As more advanced automation capabilities are being developed in Air Traffic Control (ATC), most automated systems still need human supervision. Generally speaking, automation operates as intended for scenarios it was designed to handle but requires human intervention to cope with scenarios unanticipated in the design. The success of a manual intervention depends on how much the human supervisor was (cognitively) involved in overseeing the automation's performance (Endsley, 2017). Literature and empirical evidence suggest, however, that supervisory control can cause severe vigilance and complacency problems, out-of-the-loop situation awareness and transient workload peaks (Bainbridge (1983); Endsley and Kaber (1999)).

Research performed under the umbrellas of the Next Generation Air Transportation System (NextGen) and Single European Sky ATM Research (SESAR) underline the importance of maintaining high situation awareness (SA) and task engagement in highly automated ATC environments (Chiappe et al. (2012)). But how can we ensure that controllers maintain high situation awareness (SA) while supervising a fully automated ATC system for a prolonged period of time, without being subjected to complacency and boredom issues?

In this study an interactive task assistant was developed, the Task Engagement Tool (TET). During periods of supervisory control, at regular time intervals the TET asked questions, targeting the three levels of SA: perception, comprehension and projection. The purpose of the tool was not to measure the SA of the controller during the supervisory control task, but rather to encourage the operator engagement while automation controlled the air space. In other words, the SA queries that are commonly used to probe an operator's SA are applied here to *direct* the operator's attention to task-relevant information, engaging her in the task at hand and in doing so prepare for a possible manual takeover when automation fails. The prototype was tested in a small-scale human-in-the-loop experiment to evaluate its capability to increase task engagement, reduce transients workload peaks and improve the manual control performances after an automation failure.

The paper first discusses the design rationale of the task engagement tool in Section 2. The experimental design and results are described in Sections 3 and 4. The paper ends with a Discussion and Conclusions in the final sections.

2. TASK ENGAGEMENT TOOL

2.1 Overview

The TET is a secondary dialog next to the main electronic radar screen, see Figure 1. It shows a closed question (that is, a question that can only be answered with 'yes' or 'no') that a controller is expected to answer within a limited time frame, represented by a countdown time bar. It also provides feedback on the number of correctly and incorrectly answered questions, as a way to motivate

[1] E-mail: c.borst@tudelft.nl

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2022.10.224

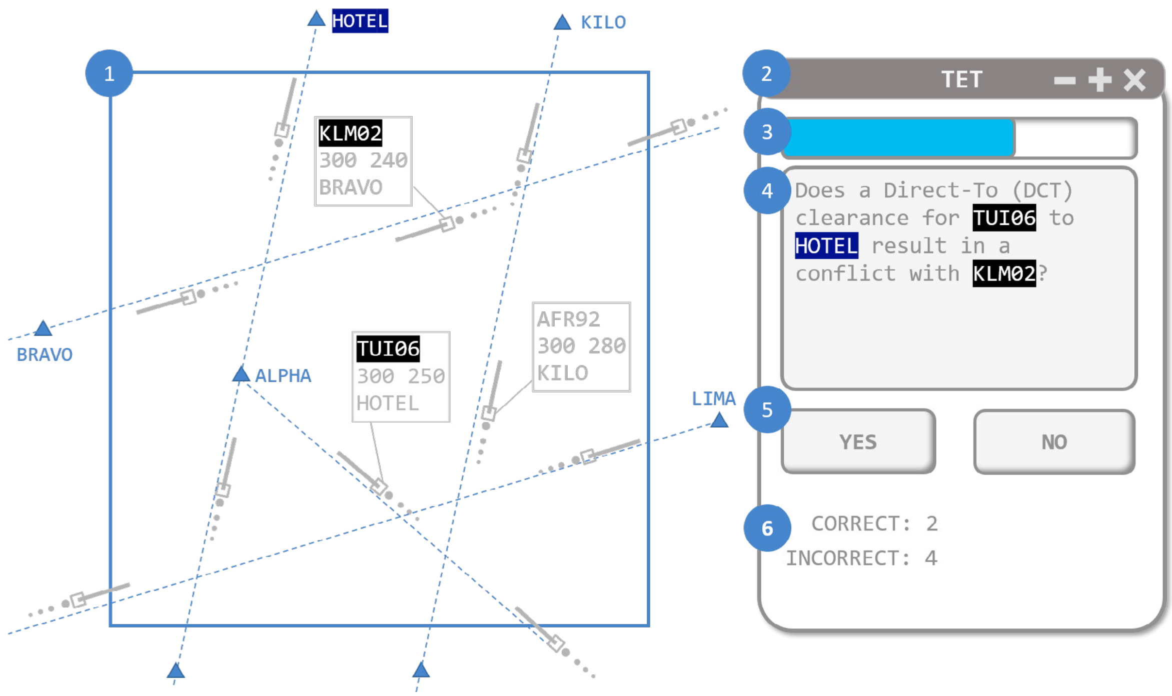

Fig. 1. Task Engagement Tool (right), a dialog next to the electronic radar screen (left, ①). The TET dialog (②) shows a countdown time bar (③), the task-related SA question (④), response buttons (⑤) and response feedback (⑥).

the controller to perform well. In essence, the TET is a *secondary* task with the purpose to mitigate boredom, increase vigilance and improve SA while supervising the fully-automated sector, which remains the *primary* task. It is expected that the TET lowers controller transient workload peaks and improves performance when she is required to assume manual control when automation fails.

2.2 ATC tasks

In this study, upper area control (UAC) is chosen as the system boundary. Compared to lower airspace, UAC has less diverse activities and could perhaps be the first part of the airspace to see full automation. From a task analysis of en-route ATC and literature in controller 'best practices' (Seamster et al. (1993); Rantanen and Nunes (2005)), at least five basic control tasks can be identified:

(1) *Maintain SA*: controllers should maintain a mental picture of the traffic, followed by continuous projection into possible future states.
(2) *Maintain orderly and efficient traffic flows*: aircraft should fly efficient routes towards the sector exit points, avoiding unnecessary additional flown track miles, minimizing time delays and maintaining an orderly structure (which makes it easier to monitor).
(3) *Detect and resolve conflicts*: when aircraft are expected to violate each other's cylindrical protected zone (radius of 5 NM horizontally and 2,000 ft vertically), they are said to be in conflict and will experience a 'loss of separation' in the near future. Controllers should detect whether conflicts are present in the sector and resolve those conflicts by issuing altitude, heading and/or speed clearances. For conflict resolution, controllers typically select resolutions that requires the least monitoring and coordination, while minimizing the number of aircraft to move and additional track miles flown.
(4) *Conformance monitoring*: controllers should check if pilots adhere to the issued clearances, by assuring all aircraft to reach their cleared target states.
(5) *Coordination with adjacent sectors*: welcoming aircraft entering the sector and transferring aircraft to neighboring sectors.

When an automated system takes over these tasks, it can be argued that the tasks of the human controller will not decrease, but increase with an additional sixth task, namely *monitoring the performance of automation.* This is especially the case when human controllers still bear the ultimate responsibility over the safety of operations. Current forms of automation are not reliable enough to safely handle all possible situation, warranting the need for human supervision and intervention.

As articulated by Bainbridge (1983) and many other studies, working alongside a highly-automated system that works well over a prolonged period of time can make human operators less vigilant, more complacent and bored. As such, controllers may fail to perform their supervisory control duties and experience difficulties in taking back control from automation when it fails. In terms of the above-mentioned ATC tasks, controllers might fail to perform task #1 above in full automation. TET aims to support this task, building a mental picture of traffic, and in doing so is hypothesized to yield sufficient vigilance, in turn enabling a swift manual takeover.

2.3 SA questions

According to Endsley and Rodgers (1994), SA comprises three levels: 1) perception, 2) comprehension and 3) projection. SA Level 1 (perception) corresponds to recognizing the status, attributes, and the dynamics of components in the environment. In terms of ATC, controllers must perceive all aircraft and their attributes in their sector, such as aircraft ID, airspeed, altitude, heading, location of waypoints, etc. The majority of this information can be retrieved directly from the radar screen and flight labels.

SA Level 2 (comprehension) is based on the integration of SA 1 elements, which forms a comprehensive mental picture of the environment when put together. Level 2 elements for ATC can entail the understanding of how fast an aircraft can change its altitude, the deviation of an aircraft from its (cleared) target state, etc.

SA Level 3 (projection) requires one's ability to project and anticipate near-future actions by combining the Level 1 and Level 2 SA elements. For ATC, SA Level 3 can include anticipating airspace capacity and judging the impact of certain routing decisions on the development of potential conflicts.

The TET questions aim to cycle a controller's SA while supervising the automated ATC system. To limit its potential intrusiveness, the questions are "closed": they only require a 'yes' or 'no' answer. The questions have been inspired by earlier ATC studies (e.g., Endsley and Rodgers (1994)) that used SA questionnaires as a way to measure the awareness of controllers. Besides the example SA Level 3 question shown in Figure 1, other examples are:

- SA 1: Does the sector have [N] crossing points?
- SA 1: Is [ACID] a heavy aircraft?
- SA 2: Did [ACID] receive a heading change and a direct-to [WP] command?
- SA 2: Will [ACID] enter the sector in the next [N] minutes?
- SA 3: Does [ACID 1] need to overtake [ACID 2] to avoid a further conflict with [ACID 3]?
- SA 3: Will it take approximately [N] minutes before a conflict with [ACID 1] and [ACID 2] becomes critical?

Here, [ACID] corresponds to an aircraft ID, [WP] to a sector waypoint and [N] to a number.

3. EXPERIMENT DESIGN

3.1 Participants and tasks

Eight participants volunteered in the experiment. Due to the exploratory nature of this study and the unavailability of professional controllers, all participants were either staff members (4) or students (4) at Faculty of Aerospace Engineering, Delft University of Technology. All subjects participated in previous and very similar ATC studies and were familiar with the simulated and simplified ATC environment.

The tasks of the participants were split in a supervisory and manual control phase. In the *supervisory control phase*, participants could not control aircraft, but needed to supervise a fully automated ATC system that performed conflict detection and resolution, cleared aircraft to their exit waypoints, assumed control over aircraft entering the sector and transferred aircraft to the adjacent sectors. In addition, participants were asked to answer the questions displayed by the TET. One question was given to participants every 100 seconds after the first minute of the simulation, and remained open for 30 seconds. During this time, participants needed to read the question, formulate the answer by consulting the radar screen, and submit their response by either clicking the 'yes' or 'no' button. If a response was not given within the time window of 30 s, it counted as incorrect and a new question appeared.

Participants were told beforehand that automation could fail at some point in time. When this occurred, an alarm sounded after which participants entered the *manual control phase*. Here, the TET and automation were switched off and participants were required to take full manual control. Participants could change an aircraft's direction and/or speed by first selecting an aircraft by clicking on its radar blip, dragging its speed vector in a new direction and/or use the mouse scroll wheel to increase or decrease airspeed. Clearances were confirmed by hitting the keyboard ENTER key. As such, no altitude changes and voice R/T were used in this simplified ATC environment.

3.2 Independent Variables

In the experiment, two within-participants independent variables were manipulated:

(1) The SA level of the TET questions, having levels 'SA1-2 (low)' and 'SA2-3 (high)',
(2) Automation failure timing in the experiment, having levels 'early failure' and 'late failure'.

Whereas the SA1-2 questions were a combination of perception and comprehension, the SA2-3 questions entailed a combination of comprehension and projection.

The automation failure timing variable was introduced to investigate a possible main and/or interaction effects of an automation failure on a controller's vigilance and secondary (TET) task performance. An 'early' failure could negatively affect a participant's trust in the automated system and increase vigilance over the remaining trials, regardless of the TET (French et al. (2018)). A 'late' failure would result increased boredom, because over a prolonged period of time the automation functions perfectly. In that case, the TET might demonstrate its value more clearly.

The experiment procedure and the distribution of the two independent variables for one participant are illustrated in Figure 2. The 'early' failure timing corresponded to having a short period before experiencing an automation failure. As the supervisory control phase (indicated as AUTO in the status row in Figure 2) lasted 7 minutes, an early failure occurred after a single block of a supervisory

Failure timing	EARLY FAILURE						LATE FAILURE	
SA level	SA1-2		SA1-2		SA1-2		SA1-2	
Scenario	scene 1A (11 min)		scene 2B (11 min)		scene 3A (11 min)		scene 4B (11 min)	
Control phase	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
Duration [min]	7	4	7	4	7	4	7	4
	BREAK + questionnaire							
Failure timing					LATE FAILURE		EARLY FAILURE	
SA level	SA2-3		SA2-3		SA2-3		SA2-3	
Scenario	scene 2A (11 min)		scene 3B (11 min)		scene 1B (11 min)		scene 4A (11 min)	
Control phase	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
Duration [min]	7	4	7	4	7	4	7	4
	questionnaire							

Fig. 2. Experiment matrix for one participant; the order was counterbalanced for each participant.

control phase. The 'late' failure occurred after five consecutive supervisory control phases leading to 29 minutes. It was anticipated that this time span was sufficiently large to cause problems in maintaining sustained attention (Esterman and Rothlein (2019)). It can be seen that for the second set of experiment runs, the location of the 'early' failure has shifted behind a 'late' failure, which still resulted in 29 minutes of supervisory control for the late failure scenario, and 7 minutes of supervisory control for the early failure scenario.

3.3 Traffic scenarios and automation

The sector used for all scenarios had two variants, A and B, of a single 'base' scenario illustrated in Figure 1; they shared identical traffic routes, but one was rotated over 180 degrees in order to prevent recognition. Waypoint names and aircraft IDs were changed for each scenario variant. Each scenario contained 30 aircraft in total, of which on average 12 aircraft were present in the sector at any given moment in time. Each scenario had 10 conflicts within the 11-minute run, and the scenarios featuring an automation failure (i.e., Scenarios 1 and 4) had four conflicts after the automated system ceased to work.

To test multiple experiment runs per participant and ensure that there would be a sufficient number of activities to interact with during the manual control phase, the simulation ran twice as fast. This yielded a real-time scenario of 1,320 seconds, which lasted for 660 seconds in the simulation. This was designed in such way that a SA question could be given every 105 seconds, which yielded 6 SA questions in a run without an automation failure, and 4 SA questions in a run with an automation failure.

Regarding the automated system itself, it was chosen to script all automated actions for the sake of experimental control (i.e., to limit variability across scenarios and to maintain comparability between participants). Thus, each scenario had a fixed set of events/actions loaded for a specific set of aircraft, where events corresponded to the ATC tasks listed in Section 2.2. The scripted events were all generated by an experienced staff member who underwent a five-day area control course at the Netherlands Aerospace Center (NLR).

3.4 Control variables

The control variables (that aimed to circumvent experimental confounds) were as follows:

- All aircraft were flying on the same altitude of Flight Level 290, which resulted in a 2D control task in the horizontal plane only. The simplification ensured that results between participants would be more comparable, as they could only change heading and/or speed of aircraft when the automation failed.
- Aircraft count, routing structure and number of predefined conflicts were fixed, resulting in a fixed complexity level for each scenario.
- All aircraft were either medium- or heavy-type aircraft. The medium type had an indicated airspeed (IAS) envelope of 200-290 kts, and the heavy type had IAS envelope of 230-350 kts.
- Automation reliability was defined as the percentage of time the automation functioned perfectly, which was true for 72 minutes of the total 88 minutes experiment duration per participant. This resulted in an automation reliability of 82%.

3.5 Dependent measures

The dependent measures were:

- *TET performance*, counting the number of correct and incorrect answers and the response times.
- *Self-reported workload*, on a 0-100 scale at a fixed two-minute interval during each scenario.
- *Manual control performance*, in terms of control inputs, horizontal separation distance and the fraction of time at which the Short Term Collision Avoidance (STCA) alarm was triggered after an automation failure event.

After each experiment session, a questionnaire collected participants' feedback on the TET, see Figure 2.

3.6 Hypotheses

First, it was hypothesized that the SA1-2 questions were the quickest and easiest to answer correctly compared to the SA2-3 questions, as they required less cognitive effort (H1). However, the SA1-2 questions would also lead to a larger peak in experienced workload when switching to manual control, because these questions did not cognitively prepare participants well for a possible manual takeover. Hence, our second hypothesis was that manual control performance was worse when confronted with SA1-2 questions in the supervisory control phase (H2). Third, we hypothesized that these trends would be larger in the presence of a 'late' automation failure (H3).

4. RESULTS

4.1 TET performance

TET metrics were defined as the number of correct responses (max. 4 for each run with an automation failure) and the response time in seconds, which can be seen in Figure 3. Regarding the number of correct responses, a non-parametric Friedman test and post-hoc pairwise comparisons did not reveal a significant difference between the two TET SA levels and automation failure timings. This can be explained by the observed behaviors and participants' feedback after the experiment, in which they indicated that some of the SA2-3 questions were a bit unclear, resulting in guessing the answer. Thus, the number of correct responses cannot be considered as a good metric for assessing the participants' SA, even though participants may have correctly updated their mental model based on the correct/incorrect counter on the TET window.

Regarding the average response time for the TET SA questions, it can be observed that participants spent more time answering SA2-3 questions than SA1-2 questions, as hypothesized. Repeated measures ANOVA revealed that the TET SA level manipulation had a significant main effect ($F(1, 7) = 7.824$, $p = 0.027$), while the automation failure timing manipulation and the interaction of two manipulations did not have a significant effect.

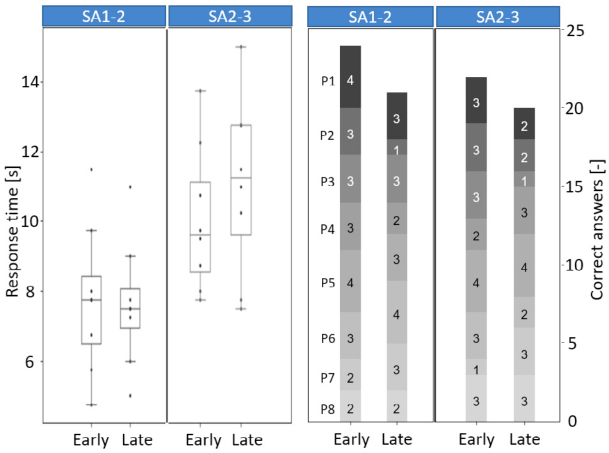

Fig. 3. TET performance in terms of response times and number of correct answers, categorized by SA level and automation failure timing.

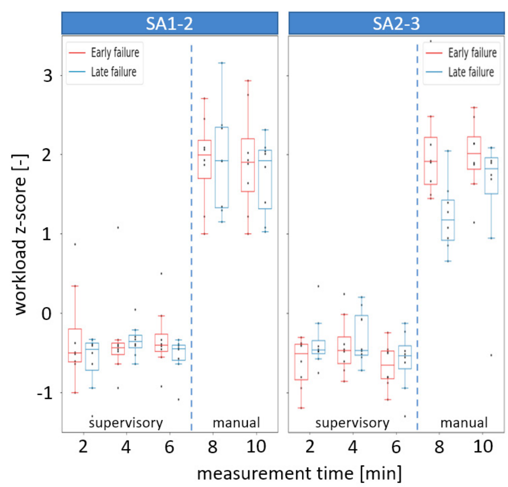

Fig. 4. Self-reported workload.

4.2 Workload

The boxplots of the z-scored workload ratings (per time interval and control phase) are shown in Figure 4. Clearly, after the automation failure the workload spikes. A repeated-measures ANOVA revealed that the SA manipulation had a significant effect ($F(1,7) = 6.784, p = 0.035$) at 8 minutes. Interestingly, the workload difference between the supervisory and manual control phases appears to be smaller when experiencing a 'late' failure while being faced with SA2-3 questions. In other words, asking SA2-3 questions during a prolonged supervisory control phase managed to close the experienced workload gap between the supervisory and manual control phases. A Friedman test confirmed a significant effect of this observation ($\chi^2(3) = 12.570, p = 0.016$) and confirms the hypothesis that the TET is most beneficial for a prolonged supervisory control phase.

4.3 Manual control performance

Considering safety, the average minimum horizontal separation distance and the time ratio the STCA was activated (measured relative to the duration of the manual control

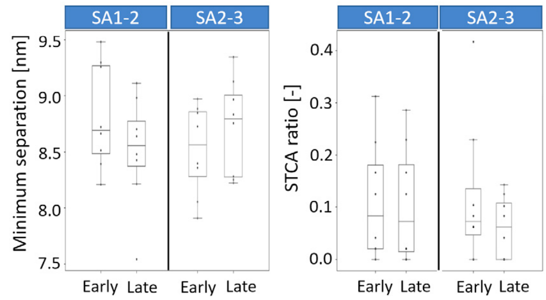

Fig. 5. Safety performance

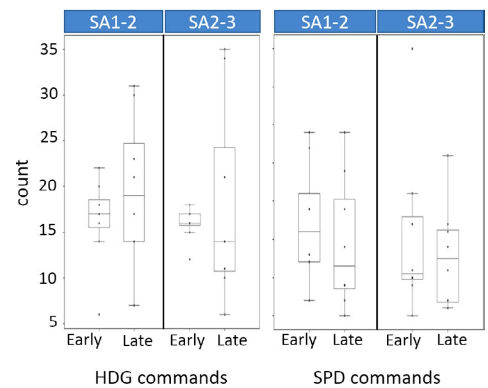

Fig. 6. Control inputs.

phase) were recorded during the manual control phases, see Figure 5. No loss of separation occurred, all minimum separation distances exceeded the 5 NM threshold. Repeated-measures ANOVA revealed that only the automation failure timing had a significant effect on the average minimum separation distance ($F(1,7) = 12.144$, $p = 0.01$).

Regarding the STCA ratio, there seems to be a weak trend (in line with the hypothesis) suggesting that with SA2-3 questions, the fraction of time the STCA alert was triggered decreased, especially for the 'late' failure. A Friedman test and consecutive pairwise comparisons (adopting a Bonferroni correction), however, did not find this trend to be significant.

Figure 6 shows the number of heading (HDG) and speed (SPD) issued after the automation failure. Both metrics (and the results of a repeated measures ANOVA) show that the SA level and automation failure timing manipulations did not have a significant effect on the number of heading and speed commands.

More interestingly, it can be observed that the number of heading commands for the 'late' failure scenario had a larger standard deviation than the 'early' failure scenario, regardless of the SA level. This result can be explained in two ways. On the one hand, it could be that some participants were more bored and 'zoned out' than other participants after being exposed to a prolonged supervisory control phase, resulting in less efficient control inputs that required more follow-up actions. On the other hand, experiencing an 'early' failure might have impacted participants' trust in automation, making them more vigilant and leading to more efficient control inputs.

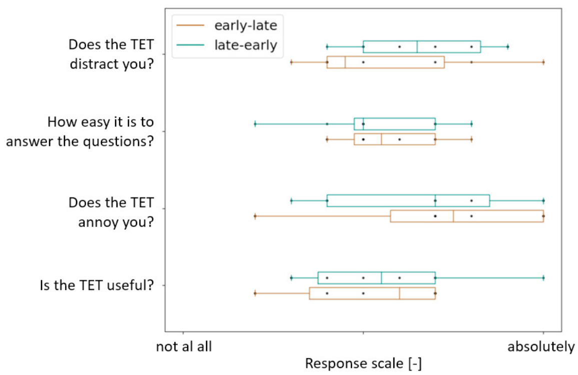

Fig. 7. Questionnaire responses.

4.4 Participant feedback: questionnaires

In general, participants found the automated ATC system useful, despite its 82% reliability. As expected, participants commented that their trust was lower when experiencing an early failure first. This, however, also led to a more positive opinion on the TET, see Figure 7. That is, participants found the TET more useful when they experienced an early failure first. This could be explained by an observation during the experiment where participants were more motivated to find answers to the TET questions. However, they also responded that the TET annoyed them more after they experienced an early failure, most likely due to difficulties in properly dividing their attention between the primary task (i.e., supervisory control) and the secondary task (i.e., answering TET prompts).

5. DISCUSSION

The TET prototype was designed to support sustained attention during the supervision of a highly automated ATC system. The goal was to cognitively engage supervisors in the automated ATC task, to better prepare them for a manual takeover in case the automation fails.

The results from a first exploratory, small-scale, human-in-the-loop experiment, featuring eight non-professional controllers in a simplified ATC environment, show that asking especially the higher-level SA questions can indeed be beneficial in lowering the workload peak between supervisory and manual control, especially for longer supervisory control phase duration. In terms of manual control performance, no significant differences were observed. It can thus be said that the TET may have increased the participants' knowledge and engagement, but that it did not result in noticeable differences in manual control performance. By design, the TET only targeted the participants' knowledge at different SA levels, but it did not allow participants to practice their hands-on manual control skills.

In hindsight, the TET as presented here may not be a good operational tool in a professional ATC setting, due to its distracting nature, not helping participants to understand the automation itself (i.e., transparency) and not provide support in practicing manual control skills. Alternatively, the TET in conjunction with the fully-automated ATC system may serve as a valuable training tool. That is, the TET can guide the attention of trainees to important low- and high-level task information, while the automated system demonstrates how to properly perform the task. As such, more research with empirical insights is needed to establish the best possible strategy on how to keep controllers engaged, vigilant and skilled in supervisory control contexts.

6. CONCLUSION

To increase task engagement and bridge the (cognitive) gap between a purely supervisory and manual control task, this paper presented the design and initial evaluation of a digital assistant, the Task Engagement Tool (TET). The TET asked task-relevant questions at different levels of situation awareness during the supervision of a fully-automated ATC system. Results from an exploratory experiment suggest that high-level SA questions were the most effective in reducing workload peaks when taking over manual control after a prolonged supervisory control period. It did not affect manual control performance after automation failed. Future studies should investigate the tool's value for training purposes and possibly explore other methods that engage controllers in supervising highly-automated systems in ways that let them practice manual control.

REFERENCES

Bainbridge, L. (1983). Ironies of Automation. *Automatica*, 19(6), 775–779.

Chiappe, D., Vu, K.P.L., and Strybel, T. (2012). Situation Awareness in the NextGen Air Traffic Management System. *Int. J. of Hum.-Comp. Int.*, 28(2), 140–151.

Endsley, M.R. and Kaber, D.B. (1999). Level of Automation Effects on Performance, Situation Awareness and Workload in a Dynamic Control Task. *Ergonomics*, 42(3), 462–492.

Endsley, M.R. (2017). From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, 59(1), 5–27.

Endsley, M.R. and Rodgers, M.D. (1994). Situation Awareness Information Requirements for En Route Air Traffic Control. Technical report, US Dept. of Transportation, Off. of Aviat. Medicine, Washington (DC).

Esterman, M. and Rothlein, D. (2019). Models of Sustained Attention. *Current Opinion in Psychology*, 29, 174–180.

French, B., Duenser, A., and Heathcote, A. (2018). Trust in Automation: A literature review. Technical report.

Rantanen, E.M. and Nunes, A. (2005). Hierarchical Conflict Detection in Air Traffic Control. *The Int. J. of Aviation Psychology*, 15(4), 339–362.

Seamster, T.L., Redding, R.E., Cannon, J.R., Ryder, J.M., and Janine, A. (1993). Cognitive Task Analysis of Expertise in Air Traffic Control. *The Int. J. of Aviation Psychology*, 3(4), 257–283.