Advances in Human Factors of Transportation, Vol. 224, 2026, 107–116
https://doi.org/10.54941/ahfe1007838

Cognitive Reinforcement for Aircrew Coordination With Autonomous Collaborative Platforms in Next-Generation Fighters

Jean-Christophe Hurault, Marianne Jarry, Grégory Froger, Anne Lise Marchand, and Colin Blättler

CREA, Centre de recherche de l'École de l'Air et de l'Espace, Ecole de l'Air et de l'Espace, F-13660 Salon-de-Provence, France

ABSTRACT

High-tempo domains such as defense, medicine, and transportation have been transformed by AI-based systems that accelerate information processing and decision support. Modern air combat reflects this evolution, with automated systems continuously generating tactical options that require aircrews to decide more frequently under time pressure. This shift is exemplified by the deployment of Autonomous Collaborative Platforms (ACPs), i.e., AI-enabled teamed drones operating alongside crewed fighters and proposing actions such as target validation. This study examines whether the processing of target-validation situations proposed by ACPs can be improved through cognitive training. Forty military aircrew members participated in a virtual reality flight mission and were assigned to either a control or a training group. Both groups completed identical pretest and posttest separated by 24 hours, while only the training group performed a 45-minute computer-based training session focused on repeated target-validation tasks. Results show that training significantly reduced error rates and response times, in line with theories of expertise acquisition. In contrast, perceived workload decreased similarly in both groups, suggesting a dissociation between objective performance gains and subjective workload. These findings support the relevance of cognitive training for preparing aircrews for future collaborative combat involving ACPs.

Keywords: Cognitive training, Next-generation fighter systems, Virtual reality simulation

INTRODUCTION

Over the past decades, high-tempo domains such as defense, finance, medicine, and transportation have undergone transformations driven by highly computerized and AI-based systems. These systems increasingly shape how information is collected, filtered, prioritized, and presented to human operators. While such developments considerably enhance data processing capabilities, they also modify the cognitive demands placed on decision makers, particularly in terms of temporal pressure, task coordination, and responsibility. In such dynamic environments, performance cannot be reduced to access to information alone. It critically depends on operators'

Received January 28, 2026; Revised April 4, 2026; Accepted April 22, 2026; Available online July 20, 2026

ability to interpret, prioritize, and act upon system-generated information and options under time pressure (Wickens et al., 2021; Cummings, 2017).

Modern air combat exemplifies this shift. The context of contemporary conflicts is increasingly characterized by multi-domain operations, requiring the simultaneous management of heterogeneous information streams such as communications, sensor data, and assets positions across domains. In response, new generation air combat programs are developing system-of-systems architectures, including the European Future Combat Air System (FCAS), which integrates a highly connected Next Generation Fighter (NGF) with remote carriers. These architectures rely on combat-cloud connectivity, high-bandwidth tactical networks, and heterogeneous sensor suites, with onboard AI intended to pre-process, filter, and prioritize information so that only decision-relevant outputs are presented to the aircrew through adapted interfaces.

At the operational level, these technological advances do not simply extend human perceptual capabilities. As automated systems generate and propose options at an increasing tempo, the number of tactical opportunities expands, requiring aircrews to assess situations and commit to decisions more frequently and more rapidly, rather than benefiting from additional deliberation time. The growing integration of AI-based onboard systems further reinforces this shift: while such systems support perception, data fusion, and option generation, they remain constrained by technical, operational, and ethical considerations, leaving critical decision-making responsibilities under human control. As a result, aircrews must not only execute mission tasks but also supervise, evaluate, and coordinate AI-generated propositions within the overall management of the combat system, contributing to substantial cognitive demands in time-critical and uncertainty-driven operational contexts.

This evolution is particularly pronounced in the context of future air combat systems, where Autonomous Collaborative Platforms (ACPs) are expected to play a central role. ACPs refer to AI-enabled teamed drones designed to operate as cooperative assets alongside crewed fighter aircraft. Acting as distributed sensing platforms and as contributors to mission execution, these systems extend the reach and responsiveness of the combat system while generating actionable propositions for human operators, notably in the form of target-validation propositions. Recent NATO-oriented analyses emphasize that, while ACPs may significantly increase operational effectiveness, they also introduce unresolved challenges related to human coordination, trust, and decision-making under time pressure (Anderson, 2025).

From a human factors perspective, the integration of ACPs does not replace existing aircrew cognitive activities but adds additional layers of cognitive demand to them. ACPs increase the frequency and complexity of coordination and evaluation processes, which must be managed alongside already established mission execution and supervision tasks, potentially leading to competition for limited cognitive resources (Wickens, 2008; Wickens et al., 2021). Aircrews are required to interpret system-generated propositions, arbitrate between multiple options, and assume responsibility for decisions whose consequences often remain ethically, legally, and operationally human-bound (e.g., lethal force employment in French doctrine).

Moreover, ACP-generated information may rely on spatial and contextual representations that require aircrews to integrate heterogeneous frames of reference into their situational awareness. Prior research on human–automation interaction suggests that such situations can be particularly susceptible to mismatches between system behavior and human cognitive capabilities, especially when operators are insufficiently prepared to understand, anticipate, and manage automated assistance (Parasuraman et al., 2000; Klein et al., 2004; Woods, 2018).

Training therefore emerges as a critical lever for supporting effective human-autonomy coordination. Theories of skill acquisition and automatization suggest that repeated exposure to structured decision situations can improve performance by increasing processing efficiency and supporting a transition from controlled, effortful processing toward more efficient retrieval-based mechanisms (Logan, 1988, 2002; Gobet & Sala, 2023).

Building on these theoretical foundations, the present study examines whether specific cognitive training can enhance aircrew performance in target-validation situations proposed by ACPs. Using an immersive virtual-reality simulation representative of next-generation fighter operations, the study assesses the effects of a short, training intervention on objective performance indicators (i.e., error rate and response time) as well as on perceived workload measured with an abbreviated NASA-TLX. The training group should show a significant reduction in error rate and response time from pretest to posttest compared to the control group, along with a significant decrease in perceived workload.

METHOD

Participants

Forty military aircrew members volunteered to participate in the study. Participants were assigned to either a control group (N = 22) or a training group (N = 18), with comparable distributions in age (Mean = 23.8 ± 2.2 years) and operational background. All participants had normal vision, prior experience with flight simulation and provided informed consent.

Stimuli and Apparatus

Simulation Task

The simulation task was implemented in a combat flight simulation environment (Mission Combat Simulator, MCS) operated in immersive virtual reality. The simulation was displayed using a Varjo XR-4 Focus Edition head-mounted display, providing high-resolution visual rendering and full cockpit immersion. Participants interacted with the simulator using a standard HOTAS configuration (joystick and throttle) while seated in a fixed cockpit seat representative of a fighter aircraft. See Figure 1.

The simulated flight scenario consisted of a 15-minutes low-altitude terrain-following route. During this flight, participants were required to maintain stable flight parameters consistent with terrain-following constraints while simultaneously responding to target-validation trials. The route and flight profile were validated in advance by a qualified military pilot to ensure operational realism and a consistent level of difficulty across participants. To control for potential route-related effects, the direction of the flight path was counterbalanced between pretest and posttest.

During the simulation, the target-validation task was presented inside the virtual cockpit using OpenKneeboard, allowing the direct integration of the task into the flight environment without interrupting the ongoing simulation.

Figure 1: Immersive flight simulation environment. View from the virtual cockpit during the low-altitude terrain-following flight, with the target-validation display integrated into the cockpit.

Target-Validation Task

The target-validation task consisted of a spatial judgment requiring participants to evaluate target propositions from ACPs. The task was implemented using OpenSesame 4, which controlled stimulus presentation, timing, and response logging.

On each trial, participants were presented with a pod-like visual display, with a defined orientation, depicting a target element and a reference element. This display was accompanied by a short declarative statement describing the spatial relationship between the target and the reference. Participants were required to determine whether the statement accurately described the spatial configuration shown.

Each trial was initiated by a brief auditory cue combined with a salient visual signal (yellow flash, 500 ms), indicating the onset of a target-validation request. The visual display and associated proposition then appeared and remained visible until a response was provided or until a maximum response time of 10 seconds. Participants responded using a binary “correct” or “incorrect” judgment via dedicated response controls.

Stimuli were generated to avoid trivial spatial configurations, and propositions were balanced to ensure an equal proportion of correct and incorrect trials. During the training phase only, feedback was provided after each response for a duration of 5 seconds. This feedback screen displayed the previous target-validation proposition, information about response correctness and proposition validity, as well as a compass aligned with the visual display to facilitate interpretation of the spatial relationship between the target and the reference. See Figure 2.

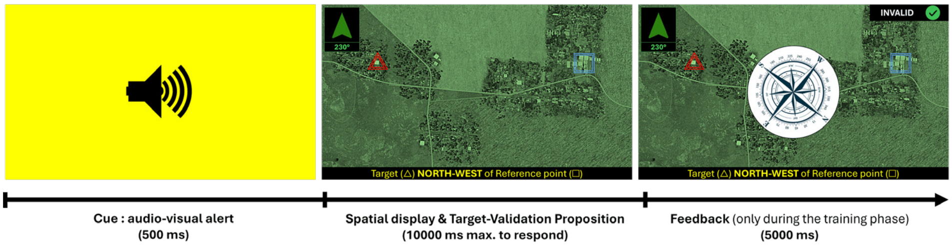

Figure 2: Temporal structure of a target-validation trial.

PROTOCOL

The experiment followed a pretest–training–posttest design and was conducted over two days. The first session included the pretest simulation phase, followed by the training phase for the experimental group only. The second session took place 24 hours later, at the same time of day, and consisted solely of the posttest simulation phase. The total experimental duration across both sessions was approximately 1 hour and 40 minutes.

Simulation Phase (Pretest and Posttest)

All participants completed a pretest and a posttest simulation; each session lasted about 30 minutes. The pretest included a 10-minute instruction and habituation phase, allowing participants to familiarize themselves with the simulator, controls, and displays. This habituation phase was reduced during the posttest.

The simulation task consisted of a 15-minute low-altitude terrain-following flight, during which the target-validation task was embedded within the flight activity. Ten target-validation trials occurred during each flight, at intervals of 90 seconds. Participants were required to maintain flight control while responding to target-validation requests. No performance feedback was provided during the simulation.

Measures collected included target-validation error rate and response time, as well as perceived workload, which was assessed at the end of each simulation session using a raw, abbreviated version of the NASA Task Load Index (NASA-TLX) covering Mental Demand, Temporal Demand, Performance, and Effort.

Training Phase

Participants assigned to the training group completed a specific cognitive training session, just after the pretest. This training was performed on a standard PC (17-inch screen, 1920x1080 px) using OpenSesame 4 and lasted approximately 45 minutes. The training task consisted exclusively of repeated target-validation trials, without any flight-related activity.

The training session comprised eight blocks of 20 trials (160 trials in total). Trials followed a faster pace compared to the simulation condition. Within each block, trials were presented consecutively without interruption. Each trial was initiated by an auditory cue and a visual flash (500 ms), followed by presentation of the spatial display. Participants had up to 10 seconds to respond, after which explicit feedback indicating response correctness was provided for 5 seconds before the next trial. Participants were allowed to take short breaks of several tens of seconds between blocks, at their discretion, to limit fatigue while maintaining engagement with the task.

Participants assigned to the control group did not perform any intermediate task and were free to leave the experimental room after the pretest simulation. All participants were explicitly instructed not to engage in any form of self-training or practice related to the task before returning for the posttest session.

RESULTS

Target-Validation Performance

Error Rate

Error rates were analysed using a mixed repeated-measures ANOVA with Test (Pretest vs. Posttest) as a within-subjects factor and Group (Control vs. Training) as a between-subjects factor.

The analysis revealed a significant main effect of Test, $F(1, 38) = 36.63$, $p < .001$, $\eta^2p = .49$, indicating an overall reduction in error rates from pretest to posttest. A significant main effect of Group was also observed, $F(1, 38) = 5.63$, $p = .023$, $\eta^2p = .13$, reflecting lower error rates overall in the training group compared to the control group. Importantly, the Test × Group interaction was significant, $F(1, 38) = 7.58$, $p = .009$, $\eta^2p = .17$, indicating that changes in performance over time differed between groups.

Post hoc comparisons using Holm correction were conducted to clarify this interaction. At pretest, no significant difference in error rate was observed between the control and training groups ($p = .57$), confirming comparable baseline performance. Within the control group, the reduction in error rate from pretest to posttest did not reach statistical significance ($p = .06$). In contrast, the training group showed a significant decrease in error rate from pretest to posttest ($p < .001$). At posttest, the training group exhibited significantly lower error rates than the control group ($p = .005$).

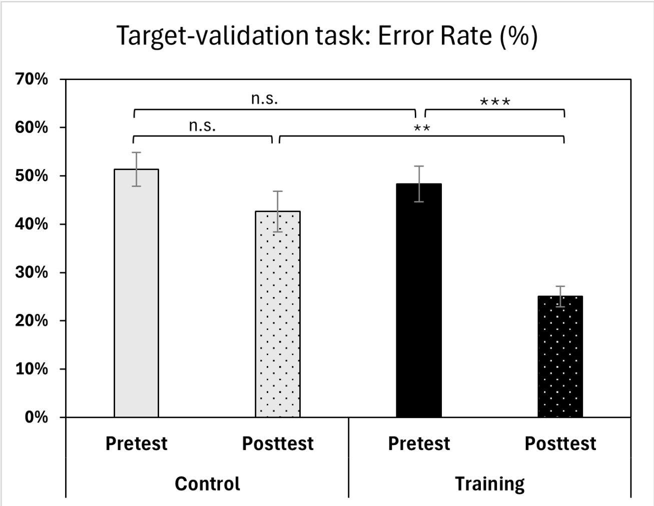

Figure 3: Mean error rates at pretest and posttest for the control and training groups. Error bars represent standard errors. Asterisks indicate significant differences.

Descriptive statistics further illustrate this pattern: while both groups showed a numerical improvement, the reduction was markedly larger in the training group (pretest: M = 0.48, SD = 0.15; posttest: M = 0.25, SD = 0.09) than in the control group (pretest: M = 0.51, SD = 0.17; posttest: M = 0.43, SD = 0.19). Overall, these results indicate a specific and robust training-related improvement in target-validation task accuracy. See Figure 3.

Response Time

Response times were analysed using the same mixed ANOVA design.

Results revealed a strong main effect of Test, $F(1, 38) = 109.17$, $p < .001$, $\eta^2p = .74$, reflecting a substantial reduction in response times from pretest to posttest across participants. A significant main effect of Group was also observed, $F(1, 38) = 6.73$, $p = .01$, $\eta^2p = .15$, indicating overall faster responses in the training group compared to the control group. The Test × Group interaction was significant, $F(1, 38) = 10.75$, $p = .002$, $\eta^2p = .22$, demonstrating that the magnitude of improvement differed between groups.

Post hoc analyses showed no significant difference between groups at pretest ($p = .29$), confirming equivalent initial response times. Both groups exhibited significant reductions in response time from pretest to posttest (control group: $p < .001$; training group: $p < .001$). However, the decrease was significantly larger in the training group. At posttest, the training group responded significantly faster than the control group ($p = .008$).

Descriptively, response times decreased from M = 7521 ms (SD = 810) to M = 5956 ms (SD = 1026) in the training group, whereas the control group showed a smaller reduction, from M = 7888 ms (SD = 838) to M = 7070 ms (SD = 1139). These findings indicate that cognitive training led to more efficient processing during the target-validation task, beyond general practice effects. See Figure 4.

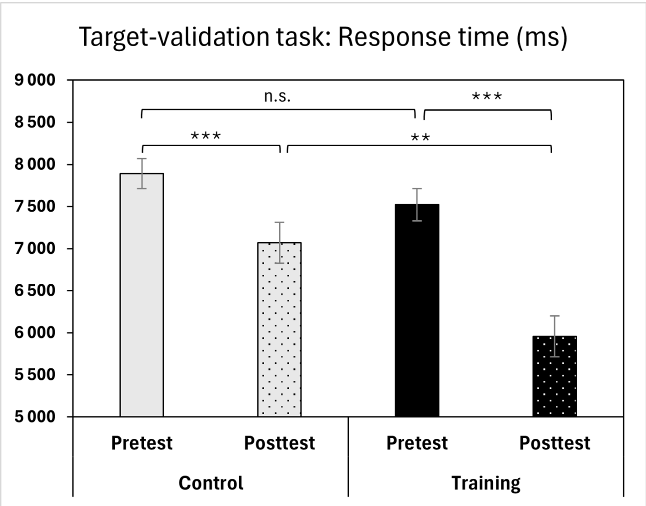

Figure 4: Mean response times at pretest and posttest for the control and training groups. Error bars represent standard errors. Asterisks indicate significant differences.

Perceived workload (NASA-TLX)

Perceived workload scores were analysed using a mixed repeated-measures ANOVA with Test (Pretest vs. Posttest) and Group (Control vs. Training).

The analysis revealed a significant main effect of Test, $F(1, 38) = 21.50$, $p < .001$, $\eta^2p = .36$, indicating a general decrease in perceived workload from pretest to posttest. In contrast, neither the main effect of Group, $F(1, 38) = 0.68$, $p = .413$, $\eta^2p = .02$, nor the Test × Group interaction, $F(1, 38) = 0.33$, $p = .570$, $\eta^2p = .01$, reached statistical significance.

Descriptive statistics indicate that perceived workload decreased in both groups between pretest and posttest, but without a differential effect attributable to the cognitive training intervention (Control: Pretest M = 14.4, Posttest M = 13.2; Training: Pretest M = 14, Posttest M = 12.5). See Figure 5.

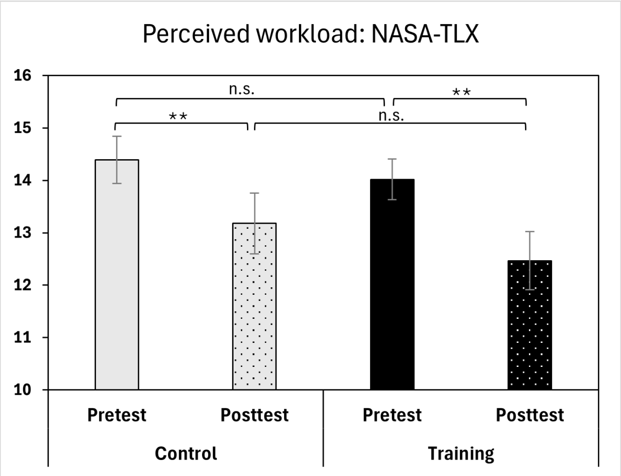

Figure 5: Mean perceived workload scores at pretest and posttest for the control and training groups. Error bars represent standard errors. Asterisks indicate significant differences.

DISCUSSION

The results on objective performance demonstrate that specific cognitive training led to significant improvements in spatial target-validation performance, as reflected by both reduced error rates and faster response times. This pattern is consistent with theories of skill acquisition and automatization (Logan, 1988, 2002; Gobet & Sala, 2023), which describe a progression toward more efficient processing through repeated exposure to structured situations. Importantly, these gains significantly exceeded the test–retest improvements observed in the control group and were consistently supported by post hoc analyses. This suggests that participants in the training group developed more efficient strategies for processing Autonomous Collaborative Platforms (ACPs) generated target-validation proposals.

In contrast, subjective measures indicated a general decrease in perceived workload from pretest to posttest for both groups, with no training-specific effect. This reduction is likely attributable to habituation to the simulation environment and task structure, a common effect in repeated-exposure paradigms. Taken together, these results indicate a dissociation between objective performance gains and subjective workload assessment: while training selectively improved performance, perceived workload decreased similarly for trained and untrained participants.

Several interpretations may account for this pattern. Subjective workload measures such as the NASA-TLX may be insufficiently task-specific to capture training-related improvements in a fast-paced, multitask context. In the present paradigm, perceived workload likely reflects a global appraisal of mission demands rather than the efficiency of the target-validation process itself. Consistent with theories of skill acquisition, performance gains may thus emerge at the task level without translating into a differentiated subjective experience of workload, particularly under sustained time pressure and operational uncertainty.

From an operational perspective, these findings indicate that it is feasible to design specific cognitive training programs aimed at improving task coordination between aircrews and ACPs. Even with a relatively short and focused intervention, training enhanced the efficiency with which aircrews processed and evaluated ACP-generated propositions. Such results support the relevance of short, ground-based training solutions to prepare operators for collaborative combat situations characterized by frequent decision cycles and high temporal constraints. At the same time, they highlight the importance of combining multiple subjective and objective measures of cognitive cost, as recommended in recent workload assessment frameworks (e.g., Longo, 2022), to capture complementary facets of operator state.

Importantly, the present findings do not address the intrinsic capabilities or limits of AI systems themselves, but rather the evolving role of the human operator within AI-supported decision processes. The absence of a training-specific effect on perceived workload suggests that, even when performance improves, operators may continue to experience such situations as cognitively demanding. This study also highlights the interest in developing additional objective indicators of cognitive workload (for example, through eye-tracking, EEG…), potentially allowing a more detailed real-time monitoring of the operator's mental state at-beyond the traditional subjective measures.

CONCLUSION

In conclusion, this study demonstrates that specific cognitive training can significantly enhance aircrew performance in the target-validation task proposed by Autonomous Collaborative Platforms, even over a short training duration. By reducing error rates and response times, such training appears to improve the efficiency of human–autonomy teaming in time-critical operational contexts. Furthermore, subjective measures of workload, in this context, may have limitations that could be overcome by developing objective indicators.

These findings support the integration of cognitive reinforcement strategies as a complementary component of future human–autonomy teaming, contributing to aircrew preparation for next-generation collaborative combat. Future research should examine how these effects evolve with longer training periods, different levels of automation, and richer combinations of subjective and objective measures of cognitive cost, to better prepare aircrews and inform the design of autonomous systems aligned with human cognitive constraints.

DECLARATION OF INTEREST STATEMENT

The authors report that there are no competing interests to declare.

ACKNOWLEDGMENT

Many thanks to all the participants for their time and cooperation.

FUNDING

This work was supported by the Agence de l'Innovation de Défense under Grant [number 2023_65_0085].

REFERENCES

Anderson, Kevin. (May, 2025). The Missing Pieces of NATO's Autonomous Collaborative Platform Strategy. Journal of the Joint Air Power Competence Centre Website: https://www.japcc.org/articles/the-missing-pieces-of-natos-autonomous-collaborative-platform-strategy/

Cummings, M. L. (2017). Automation bias in intelligent time critical decision support systems. In Decision making in aviation (pp. 289-294). Routledge. https://doi.org/10.4324/9781315095080-17

Gobet, F., & Sala, G. (2023). Cognitive training: A field in search of a phenomenon. Perspectives on Psychological Science, 18(1), 125-141. https://doi.org/10.1177/17456916221091830

Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten challenges for making automation a team player. *IEEE Intelligent Systems*, 19(6), 91–95. https://doi.org/10.1109/MIS.2004.74

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492. https://doi.org/10.1037/0033-295X.95.4.492

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological review*, *109*(2), 376. https://doi.org/10.1037/0033-295X.109.2.376

Longo, L., Wickens, C. D., Hancock, G., & Hancock, P. A. (2022). Human mental workload: A survey and a novel inclusive definition. Frontiers in psychology, 13, 883321. https://doi.org/10.3389/fpsyg.2022.883321

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A*, 30(3), 286–297. https://doi.org/10.1109/3468.844354

Wickens, C. D. (2008). Multiple resources and mental workload. Human factors, 50(3), 449-455. https://doi.org/10.1518/001872008x288394

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). Engineering psychology and human performance. Routledge. https://doi.org/10.4324/9781003177616

Woods, D. D. (2018). Decomposing automation: Apparent simplicity, real complexity. In Automation and human performance (pp. 3–17). CRC Press. https://doi.org/10.1201/9781315137957-1