Human Factors in Robots, Drones and Unmanned Systems, Vol. 138, 2024, 105–114
https://doi.org/10.54941/ahfe1005013

Identification of Management and Supervision Critical Tasks for a Multi-Effector Automated System. An Applied Cognitive Approach

Benjamin Coulomb[1], Julien Donnot[2], Aurélie Klein[3], Jean-marc Andre[4], and Françoise Darses[5]

[1]MBDA, Le plessis-robinson, 92350, France
[2]French Air Warfare Center, Mont-de-Marsan, 40000, France
[3]MBDA, Le plessis-robinson, 92350, France
[4]IMS, Bordeaux, 33000, France
[5]French Armed Forces Biomedical Research Institute, Brétigny-sur-Orge, 91220, France

ABSTRACT

Recent advances in artificial intelligence have opened up new possibilities for automation, particularly in dynamic, high-risk environments such as fighter aircraft operations. Fighter pilots' activities will change considerably as a consequence, notably when they have to manage and supervise the post-release phase of future, automated multi-effector systems. These systems will be able to adjust their own actions, and will therefore become the pilot's partner, capable of interacting like real teammates. An exploratory study carried out within a French fighter pilot squadron identified a set of tasks comprising this new activity, and linked to the use of these future multi-effector systems. The findings highlight two key issues: (i) the need to measure the importance of each task in relation to the overall mission; and (ii) the need to quantify the reduction in mental effort enabled by the automation of these tasks. The study presented here focused on identifying the tasks assessed as critical by fighter pilots when using these multi-effector systems in the post-release phase. We also aimed to determine which tasks would benefit from automation, in order to reduce the cognitive cost without compromising operational performance. To this end, we immersed 21 pilots in a simulated air operation that included the tasks inherent in managing the multi-effector system, post-release. The pilots assessed the criticality of each task for mission success, and estimated the mental effort required for each operating mode (which increasingly incorporate automation). The aim was to reveal potential cognitive benefits associated with automating each task. Operating modes were adapted from cooperation modes. Our results identified significant differences in both criticality and mental effort, depending on the task and the level of automation. Our findings make it possible to prioritize a set of tasks linked to the management of the firing plan, for the integration of automation into new weapon systems. The present research underlines the imperative to understand both the cognitive and operational needs of pilots in the context of this technological evolution, to ensure effective cooperation between the human and the system in high-risk environments.

Keywords: Fighter pilots, Swarm, Management, Supervision, Mental effort, Criticality

INTRODUCTION

Advances in artificial intelligence are enabling significant progress in automated systems, notably intelligent autonomous systems, in the military field. These technologies are being exploited by MBDA, which is developing new end-effectors with high levels of automation and autonomy. Their functions will be heterogeneous, and they will be able to operate in a restricted, coordinated group (homogeneous or not), called a 'swarm', defined using the terminology advocated by the Joint Center for Concepts, Doctrines and Experimentation. These effectors can be reconfigured in the post-release phase, which begins once the swarm has left the aircraft, and ends when the pilot enters the return-to-base phase. Managing and supervising this swarm is a new activity for the pilot. The swarm is expected to be equipped with automated, autonomous capabilities to guarantee optimum operational functioning for the pilot, and preserve his or her performance in survival-critical tasks (e.g. radar surveillance and navigation).

In order to characterize this new activity, MBDA, in close collaboration with the Centre d'Expertise Aérienne Militaire (CEAM), conducted an exploratory study with operational fighter pilots. Based on a tangible simulation method (Hauret et al., 2016), we identified eight piloting tasks that are likely to be carried out during the post-release phase. These tasks are as follows:

- Change of coordinates (Coor);
- Adding a target to the firing plan (TgtSupp);
- Swarm trajectory change (ModTraj);
- Create a Fireplan;
- Modify the function of an effector (ModFct);
- Add a new threat to the weapon system (ThtSupp);
- Swarm supervision under threat (Sup);
- ReAtt decision.

Maximizing performance requires a detailed study of each task, starting with those that are crucial to the pilot. In order to focus our research efforts, we needed to prioritize these eight tasks according to precise criteria. Following an exploratory study with a sample of pilots, we identified the following issues.

Firstly, the pilots felt that these eight tasks were not all equally critical to mission continuity. Criticality is defined as "the importance of getting the task done correctly in terms of its negative effects should problems occur" (Yanco & Drury, 2002, p. 7). The criticality of a task is likely to increase cognitive load, by altering attentional focus (Hanson, 2015). Excessive criticality could reduce performance and endanger the pilot.

Secondly, they felt that the mental effort required to carry out these tasks would be greater or lesser, depending on the interactions between the swarm and the pilot. Failure to take these interactions into account can create automation hazards (Bainbridge, 1983; Hoc et al., 2006), which is why we understand pilot-swarm interactions as a cooperative activity. In particular, we draw on work on human-machine cooperation by Hoc and colleagues

(Hoc, 2004; Hoc et al., 2006). We define pilot-swarm interaction as a cooperative activity that takes place in four modes:

- Perceptual mode (PR), which aims to stimulate the pilot's sensorimotor loop so that he or she can adapt his or her reaction.
- Mutual control (MC) mode. Here, the automated system performs the same task as the operator, in parallel, and it can guide, limit or warn him or her to avoid errors (comparable to the lane-keeping assistance system in automobiles).
- Function delegation (FD) mode, where the driver delegates all or part of the task to the system. However, the parameters required to perform the task are supplied by the operator, who also authorizes the system's actions.
- Fully automated mode (AT). In this mode, task and parameter settings are assigned to the system. The operator plays no role in carrying it out.

These cooperation modes, which reflect increasing automation, were initially developed for the task of maintaining a car in its lane. Navarro et al. (2021) applied the classification to the world of aeronautics, and automated three MATB-II microworld tasks according to the four modes. The latter authors studied the impact of expertise on cognitive load as a function of the cooperation mode. Their study demonstrated a significant decrease in perceived cognitive load when using the PR, MC, and FD cooperation modes among a sample of civilian commercial pilots. The reduction in workload was greater in FD mode than in the other modes. Similarly, expert pilots seem to favor FD mode, as it enables them to apply their expertise, while maximizing performance. The results of these earlier studies support what the sample of pilots in our exploratory study reported, and show that different modes of cooperation could have an impact on the mental effort required to complete a task.

Objective and Assumptions

The aim of the present study is to prioritize tasks for fighter crews, and, therefore, priorities for a subsequent ergonomic study. A *priority task* is defined according to two, complementary criteria: (*i*) criticality; and (*ii*) the cognitive gain provided by the cooperative mode. The ergonomic study must focus on maximizing the performance of the human-system team. In this context, we formulate several hypotheses:

- H1. Tasks related to the management of the firing plan (Coor, TgtSupp, ModTraj) determine whether the end-effectors will reach the target, which is the decisive factor in mission success. We assume that the criticality of these tasks will be higher than that of the other task categories.
- H2. The distribution of activities between the pilot and the system varies according to the mode of cooperation. The four modes increasingly integrate the system's role in task completion. The difficulty of task completion for the pilot is therefore likely to be reduced. As mental effort is correlated with task difficulty, we assume that it will decrease with increasing automation.

- H3. Coor, TgtSupp, ModTraj, FirePlan and ReAtt tasks are complex, and require a great deal of interaction with the system. Automation reduces the number of interactions, and thus the complexity of the task. As a result, we hypothesize that the gain in mental effort will be greater in these tasks.

Method and Materials

Our study included 17 fighter pilots, and four weapon system officer navigators from the French Air and Space Force. The study population was composed of 15 patrol leaders (PL), five sub-patrol leaders (SPL), and one operational pilot (OP). Participants were considered highly experienced, having carried out an average of 3.7±2.8 overseas operations (OPEX) on Rafale aircraft, and 1.4±2.3 OPEX on other military aircraft. All participants took part in the study on a voluntary basis.

The test took place remotely, and was divided into two parts. Firstly, participants took part in a briefing that lasted around 15 minutes. They were informed of the swarm's technical capabilities, the details of the mission to be carried out, and the elements needed to prepare for it.

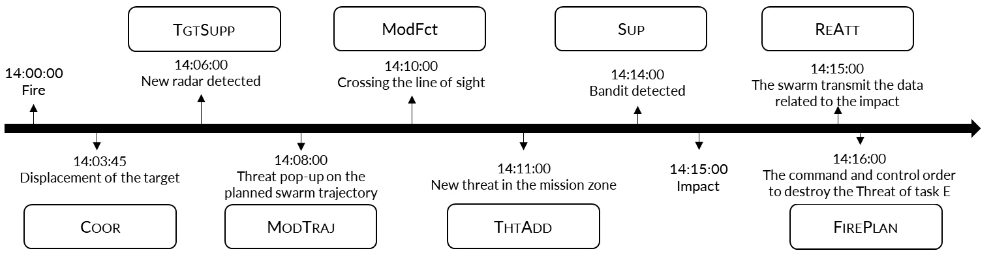

Figure 1: Mission scenario containing the eight events.

Secondly, during the experimental phase (approx. 45 minutes), participants were immersed in a scenario that simulated the post-release phase of a mission using the swarm. The objective of this mission was to neutralize an enemy air defense system. This required the swarm to be release at a great distance from the target, meaning that its flight time was very long (15 min). During this extended phase, pilots were confronted with numerous events that required them to carry out the above-mentioned tasks. The scenario was developed with the support of a test pilot from CEAM, and two operational experts from MBDA (former fighter pilots). The participation of the latter enabled us to integrate realistic and probable events (e.g. target displacement). Each event preceded the completion of one of the tasks. Following each event, participant were asked to assess the criticality of the task for mission success. Then, for the same task, the four modes of swarm operation (one for each mode of cooperation) were described to the participant. For each mode of operation, the participant was asked to imagine him or herself performing the task, in order to assess the mental effort he or she felt they would have to commit to achieve it.

Mental effort was measured using the 10cm Visual Analogue Scale (VAS), rather than subjective tools such as the SWAT or NASA-TLX (Hart & Staveland, 1988). This was because we were not able to address all of the components of cognitive load (time pressure, frustration, stress, etc.) given the prospective context in which we were working. Moreover, the use of these scales would have considerably increased the duration of the tests. The measurement of mental effort was repeated 32 times, making it necessary to identify a lightweight, easily-comprehensible measuring device. Participants were asked to evaluate the mental effort required to perform the task on a scale running from 0–100, where 0 corresponded to *None of your cognitive capacities are allocated to this task*, and 100 corresponded to *All of your cognitive capacities are allocated to this task*. To provide a common frame of reference for all participants, they were presented with an example of a task requiring maximum cognitive capacity. Criticality was measured using the same type of VAS. Pilots were asked to rate the criticality of the task for mission continuity on a scale ranging from 0–100, where 0 corresponded to *Successful completion of the task is not a prerequisite for mission success*, and 100 corresponded to *Successful completion of the task is a prerequisite for mission success*.

Pilots could voluntarily justify each of their assessments using a dedicated space under each VAS.

Data Processing

To test the third hypothesis, mental effort was analyzed in terms of gain. In perceptual mode (PR), the system intervenes the least in task performance, and we logically considered it as a baseline. Consequently, we were able to compare the gain obtained by automating the system with the manual mode, and compare the use of the same mode between tasks.

Data analysis used Matlab® (2021a) software. We applied the Shapiro–Wilk test (the *ktest* function) to check the normality of the data. This found that criticality and gain data did not follow the normal distribution ($p<0.05$). We tested the homogeneity of variances using Bartlett's test (the *p_bartlett* function). We observed a significant difference between criticality variances for the task variable ($p<0.001$). We also observed a significant difference between gain variances for task ($p<0.05$), and mode ($p<0.001$) variables. We applied a Mauchly test (the *mauchly* function) to check the sphericity of the data. This found that the data collected were not spherical ($p<0.01$). For criticality and mental effort data, the test revealed ε of 0.50 and 0.25, respectively. Consequently, data were analyzed with a repeated-measures ANOVA, applying the Greenhouse–Geisser epsilon correction (the *ranova* function). If the ANOVA showed a main effect or interaction, we used Tukey's *post hoc* test (the *multcompare* function), and Cohen's test (the *cohend* function) to assess effect size.

The corpus of written responses was analyzed in relation to the question of criticality, in order to identify factors likely to affect it. For mental effort, an analysis of the verbal corpus enabled us to identify the strengths and weaknesses of each mode of cooperation in each task.

RESULTS

Criticality assessment for each task. The results of the criticality assessment for each task are shown in Figure 2. There was a significant difference in criticality between tasks ($p<0.001$, $F_{(3)}=6.36$). The relative criticalities between tasks were then examined using a Tukey *post hoc* test. Only the ModFct task was judged less critical than firing plan tasks (Coor, TgtSupp, ModTraj, and ReAtt). The ThtSupp task was judged less critical than the Coor task.

These results do not confirm our first hypothesis, namely that criticality would be higher for the Coor, TgtSupp and ModTraj tasks. In fact, the task of modifying the function of an effector (ModFct) is judged less critical than most of the other tasks. We observe lower variability (confirmed by Bartlett's test) for criticality, particularly for the Coor, TgtSupp, ModTraj and REATT tasks, where there appears to be more agreement. These results support our hypothesis, but do not allow us to validate it.

Evaluation of mental effort as a function of cooperation mode. Figure 3 shows the results of the analysis of the mental effort evaluation as a function of cooperation mode. There was a significant difference in mental effort according to mode ($p<0.01$, $F_{(14)}=4.08$), and a significant decrease (Tukey's *post hoc* test) in mental effort in modes incorporating automation.

Our hypothesis concerning the reduction of mental effort in cooperation modes incorporating automation is therefore confirmed.

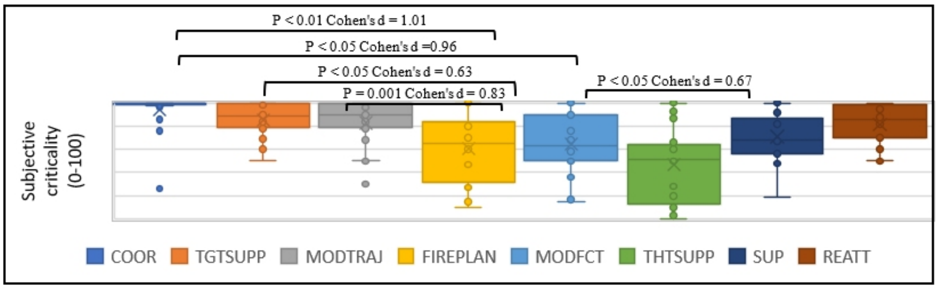

Figure 2: Criticality assessment as a function of task.

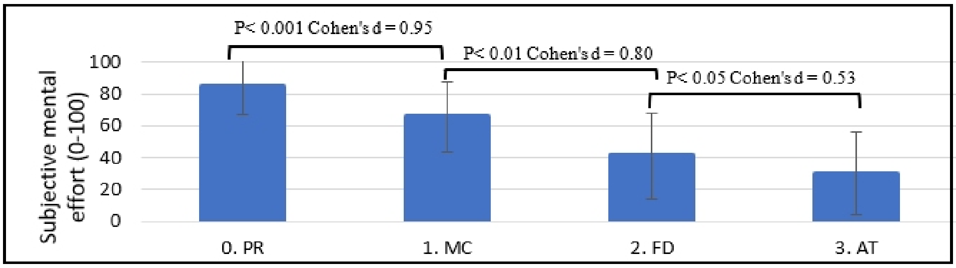

Figure 3: Mental effort as a function of cooperation mode.

Evaluation of gain in mental effort as a function of cooperation mode. With regard to gain (Figure 4), we found a simple main effect for mode of cooperation. The ANOVA revealed an interaction between *Mode* and *Task* ($p<0.001$, $F_{(3)}=44$). We also found a significant increase in gain between the following modes: (*i*) MC and FD for the tasks Coor ($p<0.001$, Cohen's $d=1.49$), TgtSupp ($p<0.01$, Cohen's $d=1.31$), ModTraj ($p<0.001$, Cohen's $d=1.50$), ThtSupp ($p<0.01$, Cohen's $d=0.17$), and ReAtt ($p<0.001$, Cohen's $d=1.81$); and (ii) between FD and AT modes for ModFct ($p<0.05$, Cohen's $d=0.85$), and ThtSupp ($p<0.05$, Cohen's $d=0.28$). When tasks were compared with each other for each mode, the only significant difference was between Coor and FirePlan for FD mode ($p<0.05$, Cohen's $d=1.02$).

These results do not allow us to validate the hypothesis of greater gains in Coor, TgtSupp, ModTraj and FirePlan tasks. In fact, when we compare tasks for the same mode, only one difference emerges—between FirePlan and Coor for the FD mode. We also observe that gains in mental effort are lower in FirePlan and ThtSupp tasks (Figure 4). These results show that FD mode offers the best gain in mental effort, as we observe a significant difference in gains between MC and FD modes in the majority of tasks, with large effect sizes. The difference between these modes lies in the extent of human intervention. In FD mode, the operator takes charge of the task, sets the necessary parameters, then adopts the best options among the proposed solutions. However, we only observe a difference between FD and AT modes in ModTraj and ThtSupp tasks. In the other tasks, AT mode does not reduce mental effort compared with FD mode.

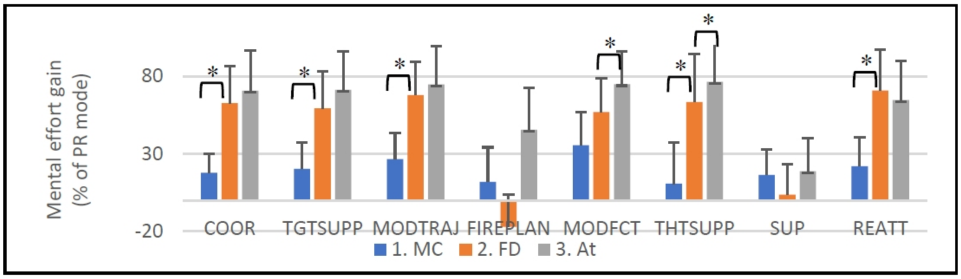

Figure 4: Gain in mental effort expressed as a percentage of the perceptual mode (PR) across modes and tasks. Brackets with asterisks represent significant differences, confirmed by Tukey's post hoc test.

Turning to justifications for their criticality and mental effort assessments, 11 out of the 21 pilots provided comments. These observations clarified the contextual factors that influence task criticality, and provided an initial assessment of the operational benefits of the different modes of cooperation. For example, qualitative elements that increased the self-direction capabilities on the swarm's final approach, together with lower target priority, could reduce the criticality of a task.

DISCUSSION

The main objective of this study was to optimize a framework for evaluating the fighter pilot's modes of cooperation with a multi-effector system. We sought to identify the priority of tasks making up the activity context. A priority task is defined as a task that is critical to mission continuity, and whose automation reduces the mental effort required to perform it.

The lower criticality of the ModFct task can be explained by the fact that it is completely new to pilots. To date, no effector can simultaneously perform several functions (jamming, decoy, neutralization). Pilots have therefore never been confronted with the task, and are able to carry out their current missions without this functionality.

Our results concerning the reduction of mental effort in different modes are in line with the literature (Stapel et al., 2019). This decrease could be due to the representation that participants construct regarding task difficulty. We know that mental effort is correlated with task difficulty (Galy et al., 2012). Thus, modes of cooperation that incorporate a high level of automation could reduce the expected difficulty of the task.

The mental gains identified in the FirePlan task do not follow the same pattern as the other tasks. However, it should be noted that the FirePlan task is more complex than the others, in that it encompasses other tasks (Coor, TgtSupp, ModTraj). The adaptation of cooperation modes for more complex tasks (such as FirePlan) seems to require improvement, particularly in FD mode. The mental gain for the Sup task is also lower than for the other tasks. Here again, this may be explained by the nature of the task, which involves supervising the swarm. For this purely cognitive task, automation resides solely in the interpretation of the information proposed by the system. As the pilot does not interact with the system, automation seems to have no effect on mental effort.

Our mental gain results are in line with the findings reported by Navarro et al. (2021). The latter authors also found a greater reduction in cognitive load in FD mode. This mode was judged to be the most suitable by pilot who justified their responses. Notably, it minimized the resources required to carry out the task, while facilitating solution verification processes. The lack of effect of AT mode can be explained by its very nature, which excludes the operator from task completion and choice validation. It implies that the swarm has the authority to make decisions, but that the pilot retains responsibility for the effectors that are dropped. The pilots who were interviewed highlighted the increased difficulty of verifying and, above all, understanding the choices made by the system. This result may be explained by the work of Deroo (2012), who demonstrated that ambiguous authority-sharing, or over-delegation to the system, creates conflict and degrades the interaction between the human and the system, with negative effects on performance. Our results can also be approached from the angle of trust (Endsley, 2023; Hoff & Bashir, 2015). A lack of trust can also degrade the effectiveness of automated mode, by forcing the driver to double-check proposed solutions.

Our results concerning mental effort need to be confirmed in future research that uses more advanced and more immersive simulations. In addition, methods will need to be developed to capture objective and subjective data measuring the operator's cognitive load.

CONCLUSION

This study paves the way for future experiments on modes of cooperation between a pilot and his or her multi-effector system. A precise definition of the tasks to be carried out is an essential prerequisite for the design of an experimental protocol for future studies. The latter should study the impact of cooperation modes on fighter pilots' ability to manage a swarm. Consequently, our future research will focus on the impact of cooperation modes on the cognitive load of the fighter pilot, when modifying the trajectory of a multi-effector system.

ACKNOWLEDGMENT

The authors wish to acknowledge that the Defense Innovation Agency and the MBDA company is funding this research project.

REFERENCES

Bainbridge, L. (1983). Ironies of automation. In Analysis, design and evaluation of man–machine systems (pp. 129–135). Elsevier.

Deroo, M. (2012). Contrôle cognitif, assistance à la conduite et coopération homme-machine : Le maintien sur une trajectoire acceptable et sécurisée [PhD Thesis]. Université Rennes 2.

Endsley, M. R. (2023). Supporting Human-AI Teams : Transparency, explainability, and situation awareness. Computers in Human Behavior, 140, 107574.

Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? International journal of psychophysiology, 83(3), 269–275.

Hanson, J. A. (2015). The effects of task criticality and target modality on a simulated battlefield search task. Old Dominion University.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index) : Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139–183). Elsevier.

Hauret, D., Donnot, J., & Van Belleghem, L. (2016). Intégration de la simulation de l'activité sur support tangible dans le processus de conception de l'interface homme-système du Rafale. Communication présentée à la 15 ème Conférence Ergo'IA.

Ho, H.-N. (2010). The relationship between levels of expertise, task difficulty, perceived self-efficacy, and mental effort investment in task performance. University of Southern California.

Hoc, J. (2004). Vers une coopération homme-machine en situation dynamique. Ergonomie, 269–284.

Hoc, J.-M. (2003). Coopération humaine et systèmes coopératifs. Ingénierie cognitive. IHM et cognition, 139–187.

Hoc, J.-M., Mars, F., Milleville-Pennel, I., Jolly, É., Netto, M., & Blosseville, J.-M. (2006). Human-machine cooperation in car driving for lateral safety : Delegation and mutual control. Le travail humain, 69(2), 153–182.

Hoff, K. A., & Bashir, M. (2015). Trust in automation : Integrating empirical evidence on factors that influence trust. Human factors, 57(3), 407–434.

Navarro, J., Allali, S., Cabrignac, N., & Cegarra, J. (2021). Impact of Pilot's Expertise on Selection, Use, Trust, and Acceptance of Automation. IEEE Transactions on Human-Machine Systems, 51(5), 432–441.

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. Transportation research part F: Traffic psychology and behaviour, 60, 590–605.

Yanco, H. A., & Drury, J. L. (2002). A taxonomy for human-robot interaction. Proceedings of the AAAI fall symposium on human-robot interaction, 111–119.