

Critical decision making with a highly automated UAVa case study

Nicolas Maille

▶ **To cite this version:**

Nicolas Maille. Critical decision making with a highly automated UAVa case study. HFES Europe Chapter Annual Meeting 2022, Apr 2022, Turin, Italy. hal-03878067

HAL Id: hal-03878067
https://hal.science/hal-03878067

Submitted on 29 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Critical decision making with a highly automated UAV – a case study

Nicolas Maille
ONERA – The French Aerospace Lab
France

Abstract

In transportation and aerospace, more automation and autonomy are continuously added to systems. The ability of human operators to effectively monitor and interact with these systems, poses significant challenges. This research focuses on critical decisions that largely rely on the system capabilities but need to be validated and made under the responsibility of the operator. In the context of unmanned combat air vehicles (UCAV), the experiment focuses on how the communication strategy of semi-autonomous systems modifies the operators' understanding of the situation and the final decision. The study has been conducted in an immersive simulator with a 30 minutes ecological military scenario where the operator had to manage a full mission, including an unplanned missile firing decision. The experiment included the use of physiological measures related to electrodermal and cardiac activities. The paper reports the results of the decision-making performances and the analyses of the physiological parameters. It appears that the communication strategy has an impact on the situation awareness of the operator, the decision taken, and the evolution of the physiological parameters.

Introduction

The development of highly automated vehicles, from autonomous cars for civilian applications to Unmanned Combat Air Vehicle (UCAV) for military operations is profoundly changing the way people interact with these systems. Although the word "autonomy" implies that systems will be able to perform actions on their own, in real-world applications, these autonomous systems must still cooperate with humans who may be responsible for effectively monitoring the behaviour of systems, directing them when needed, or acting as teammates and collaborating on decision-making. The ability of human operators to oversee and manage these systems appropriately when needed is a major challenge. Endsley (2017) wrote "*an automation conundrum exists in which as more autonomy is added to a system, and its reliability and robustness increase, the lower the situation awareness of human operators and the less likely that they will be able to take over manual control when needed*". Questions about autonomous driving and how humans adapt to taking control of these vehicles are currently under investigation (Morgan et al., 2016; Eriksson & Stanton, 2017; Morgan

In D. de Waard, S.H. Fairclough, K.A. Brookhuis, D. Manzey, L. Onnasch, A. Naumann, R. Wiczorek, F. Di Nocera, S. Röttger, and A. Toffetti (Eds.) (2022). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

et al., 2018). For example, recent studies investigate the effects of takeover signal lead time or modality on automated vehicle takeover performance (Huang & Pitts, 2022).

In the case of drones, the main issue is no longer taking manual control of the vehicle, but rather being responsible for monitoring the mission, assessing the overall situation in relation to the mission objectives, and collaborating in decision-making (Barnes & Evans, 2016). From an operational perspective, humans may soon act as managers 'on' or 'over' the decision-making loop, rather than in it (Mayer, 2015). As a result, a new context is emerging, characterized by humans managing a machine-driven decision loop. In the military domain, a strong requirement is the ability to operate in more contested air environments, which implies reduced data transmission, automated on-board analysis of raw intelligence data, and greater autonomy when navigating or tracking targets (Mayer, 2015). Increasing the level of autonomy allows for more irregular monitoring by the operator and raises the issue of "neglect time" (i.e., the length of time the system can operate autonomously before reporting back to the human) and "interaction time" (i.e., the period of time during which the system and the human communicate and define the next actions) (Olsen & Goodrich, 2003). Stress is high during the neglect period because of the uncertainty of what the system is doing and whether it will actually return a communication. Stress is also high in the interaction period because communication must be fast, efficient, and accurate (Hancock & Szalma, 2008). These communication constraints amplify the difficulties for the operator who does not have access to the continuous evolution of the situation but who may have to enter the decision loop at any moment and cooperate in highly critical and urgent decisions. While the operator has only a limited amount of time to weigh, verify and gather all critical information, such an interaction process can create a high workload and stress. It is worth pausing to reflect on the operator's ability to contribute effectively and take responsibility for the final decision.

This study focuses on the effects of the timing of communications on the decision making process. Human computer interaction studies addressing human-drone interaction generally indicate that greater transparency about the drone's behaviour helps the operator to monitor the mission (Mirri et al., 2019). Nevertheless, when communications must be sparse for operational reasons, the effect of communication choices on the operator's supervision and decision is little studied. The main contribution of this work is to evaluate in an ecological simulation how the communication strategy implemented by a UCAV impacts the final decision made by the operator and can change the stress, workload and situation awareness of this frontline operator. The use case involves a UCAV operating in a hostile environment in which a trade-off must be made between stealth (for survivability) and communications with the ground station. This work is based on two main hypotheses:

H1: An early communication strategy reduces the stress and workload of the decision-making process.

H2: An early communication strategy helps the operator to make the right decisions.

This experiment is part of a larger research project, but the results presented in this paper focus on one decision to be made during the mission. Further hypotheses on the

evaluation of global awareness and workload at the end of the mission have been defined but are not presented here.

Material and method

Participants

The study was approved by University of Aix-Marseille Ethics Committee (ref. Prop. 2018-24-05-001) and involved twenty students and junior research scientists (5 females, 15 males), aged between 20 and 38 years old (M=27.4; SD=4.9). There were all civilian employees of an aerospace laboratory. All subjects volunteered to take part in the study and gave their full informed consent before taking part in the experiment.

Task

The military operational context used to create an ecological task was based on an armed reconnaissance mission, which is one of the typical missions envisaged for future UCAVs. The main objective of the mission is to obtain detailed information on enemy activity in a given area, with a contested air space (Fig. 1). Even if the mission is not dedicated to attacking predetermined targets, the aircraft must be armed and capable of identifying threats and conducting air strikes on targets of opportunity. The mission usually involves medium-range infiltration into a contested environment, reconnaissance of the area and exfiltration from enemy territory.

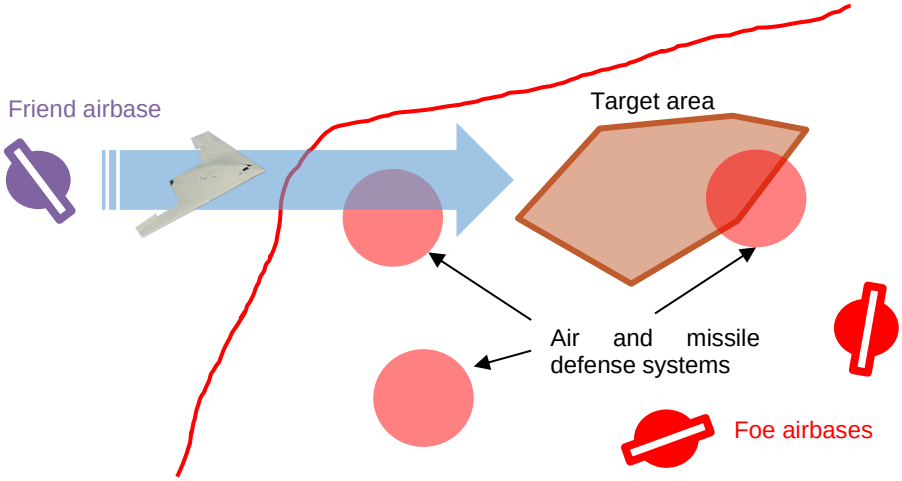


Figure 1. Armed reconnaissance mission. The red line indicates the border of the contested air space.

In this study, the simulated UCAV was equipped with terrain following capacities, air-ground missiles and highly automated features that gave it the ability to autonomously identify threats and adapt the mission if necessary. To increase survivability, communications in foe territory were severely restricted, but the UCAV took pictures of targets and threats and sent them back to the operator when a

communication point was reached. The use of the weapon to strike targets of opportunity had to comply with the rules of engagement and was under the operator's responsibility. A set of rules, in line with those used in military battlefields, was defined for this experiment. Briefly, the use of weapons was limited to enemy military targets with hostile intent and permission to fire was given to the crew only if no collateral damage was expected. Otherwise, permission to engage the target was given to a higher hierarchical level and the operator had to request permission to fire.

The participant had to monitor a complete 30 minutes mission. The initial flight plan to achieve the mission objectives was already inserted in the flight management system. This plan contained three targets to be observed by the UCAV. Pictures of these targets were to be taken and sent back to the ground station when communication between the UCAV and the ground station was allowed. Specifically, the flight plan contained sections where the UCAV communicated continuously with the ground station, updating its position, sending all available data (images) and other sections where only intermittent communication was allowed at predefined communication waypoints. Nevertheless, the initial flight plan could be modified by the UCAV if threats were encountered during the mission. Depending on the operational situation, these threats could be avoided (the UCAV moved away from the threat) or engaged (the flight plan was modified to create a missile firing opportunity on this new target). In the second case, the operator was responsible for the final firing decision, which had to be in accordance with the rules of engagement. In addition, throughout the firing window, a continuous communication channel was maintained between the UCAV and the ground station, so that all available data to support decision making was displayed to the operator. On the ground station's touch screen interface, the operator had the option to initiate or deny the attack. A radio communication system allowed direct communication with headquarters (in this case the experimenter) if approval was required by the rules of engagement before executing the action.

During the mission, two threats were identified by the UCAV and led to changes in the flight plan. Only the first, which was a missile firing opportunity, required a decision by the operator and is considered in this article. The operator had to collaborate with the UCAV and decide whether or not the target should be engaged. The process involved three distinct parts:

- Perception: extracting useful elements from the environment to understand the actual situation at that time and place on the battlefield and evaluate possible collateral damages.
- Deduction: Select the relevant rules of engagement and decide what to do. Depending on the situation, the operator could deduce that engagement was not allowed, or that engagement was only possible with the approval of higher headquarters, or that she/he could take responsibility for the missile firing.
- Action: abort the attack, contact headquarters, or validate the attack.

All participants had the same scenario and the right decision was to contact headquarters.

Experimental conditions

This study employed a 2 between-subject design (N=20). Two communication strategies (C1 and C2) were used.

C1: Early communication strategy. In this condition, the UCAV inserted new communication waypoint whenever new threats were encountered. The operator was therefore immediately informed of the change in the flight plan and knew that a firing opportunity was possible. However, the operator did not have the necessary information at that time to make the decision. This data was only provided at the beginning of the firing window.

C2: Late communication strategy. In this condition, the flight plan change was made without being communicated to the operator. The change was only sent to the operator when the firing window was started.

Thus, condition C1 favoured the transmission of new data to the operator while the second favoured the survivability of the UCAV.

Experimental device



Figure 2. UCAV flight simulator (left) and its dedicated touchscreen user interface (right).

An immersive UCAV simulator (Fig. 2) was used to run the scenario with a dedicated touchscreen interface allowing (1) the monitoring of the UCAV trajectory and the visualization of the new the flight plan when modified by the UCAV (right part of Fig. 2); (2) the visualization of the target's images sent by the UCAV (central part of Fig 2), and (3) the visualization of a continuous stream of full-motion video during the firing decision process (also on the central part of Fig 2, instead of the image management interface).

BioPac (MP150) was used to collect physiological data. Electrodermal activity (EDA) and cardiac activity (ECG) were recorded. Both raw signals were acquired at a sampling rate of 1250 Hz. The Biopac ECG100C amplifier used a band-pass filter of 35Hz and 0.5Hz. The Biopac EDA100C amplifier used with a low pass filter set at 10Hz. As both hand were used for the experiment, we used the recommended foot sites (right foot) for EDA recording (Boucsein et al., 2012).

Procedure

Once the ECG and EDA physiological acquisition systems were set up, the participants were briefed on the operational context of the mission, including the rules of engagement, and installed in the simulator. A presentation of the user interface was given and then the participants performed four training scenarios to familiarize themselves with the UCAV monitoring and the firing decision-making process. If they wished to continue with the other phases of the trial, participants signed a consent form. They then performed the 30 minutes scenario before fulfilling NASA-TLX (Hart & Staveland, 1988) (workload) and QUASA (McGuinness 2004) (situation awareness) questionnaires. Participants received a full verbal debrief. The experiment lasted approximately 2 hours.

Data analysis

A factorial independent measure design was employed. The independent variable was the experimental condition with two levels: early and late communication strategy. So 10 subjects (M=27.5; SD=5.5) managed the UCAV with an early communication strategy (C1) while the 10 others (M=27.3; SD=4.5) managed the UCAV with a late communication strategy (C2). Both electrodermal and cardiac activities were analysed thanks to the AcqKnowledge 4.1 © software.

Each participant's raw ECG data was processed using the built-in "Detect and Classify Heartbeats" function to estimate the R-wave peaks (minimum BPM 30; maximum BPM 240; R wave threshold 50% Max R peak level). A visual inspection was used to remove unreliable R peaks and related R-R intervals before calculating the mean R-R interval for the one-minute baseline and for the firing window. The metric used is then the difference, for each participant, between the mean R-R interval for the firing window and the mean R-R interval for the baseline. This normalization allows for comparison between subjects.

Each participant's raw EDA signal was visually inspected to remove parts with noisy data (foot movement) and SRC were identified thanks to the AcqKnowledge "Locate SRCs" function. For each participant, the first 500 seconds of the experiment (before the threat was detected) are used as a baseline (extract mean and SD) for a participant z-score transformation. Then, the normalized EDA allows for comparison between subjects. Two measures are used, as shown in figure 3.

- **Measure 1:** The difference between the EDA value prior to threat identification (mean value for the 60 seconds prior to detection of the new threat and communication of the flight plan change to the operator) and the higher EDA value obtained during the firing decision window.
- **Measure 2:** The largest amplitude of the skin conductance response (SCR) during the firing decision window. It measures the phasic change in electrical conductivity of the skin related to the drone request for the firing decision.

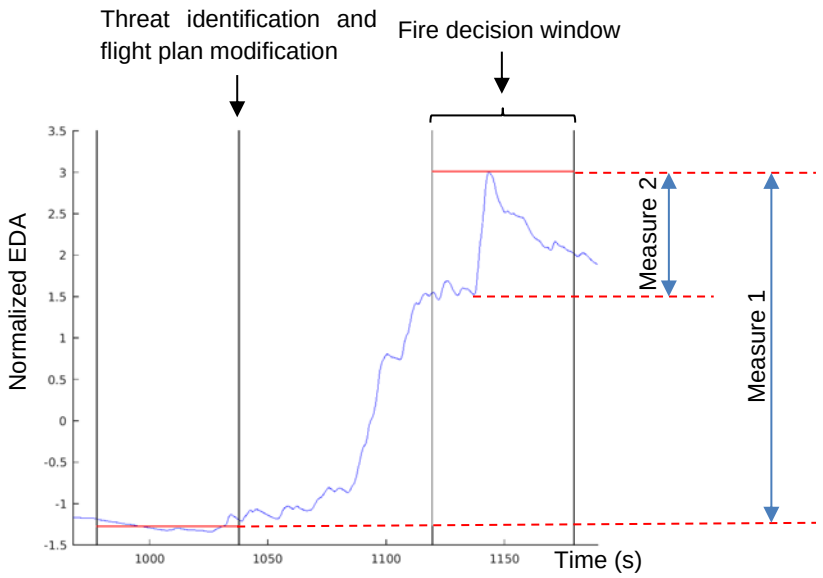


Figure 3. Measures used to characterize the galvanic skin response induced by the decision-making process.

Results

All statistical tests reported are two-tailed with alpha levels of .05. Effect sizes were determined using Cohen's d with ≥.2, ≥.5 and ≥.8 indicating small, medium and large effect sizes. In figures, the central rectangle spans the first quartile to the third quartile, the black segment inside the rectangle shows the median, the diamond gives the mean, error bars represent the "inner fence" and unfilled circles outliers.

Testing Hypothesis 1

The first hypothesis is that the early communication strategy should reduce the stress and workload associated with the decision-making process, compared to the late communication strategy. Changes in stress level and workload are assessed by physiological parameters: heart rate and skin conductance. In both experimental condition, the identification by the UCAV of a new threat and the request to the operator to validate or not an attack on this new target should increase the level of stress and workload during the firing window. In both conditions an increase in heart rate and skin conductance is expected. This first hypothesis will be confirmed if these modifications are shaped by the experimental conditions.

As already stated, the analysis of the heart rate is made on the bases of R-R intervals expressed in seconds. In both conditions, a reduction of the R-R interval is observed (Fig. 4) and corresponds to an increase of the heart rate in beat per minute. Nevertheless, the reduction of the R-R interval compared to baseline is significant only for condition C2 (Paired t-test; C1: t = 1.325, df = 9, p-value = 0.218, Cohen's d= 0.256; C2: t = 3.368, df = 9, p-value = 0.008, Cohen's d= 0.574). So, a significant

heart rate increases (in beat per minute, compare to baseline) is observed only for condition 2. The comparison of the reduction of R-R interval between both conditions is not significant (Two Sample t-test; t = 0.940, df = 18, p-value = 0.360, Cohen's *d* =0.420).

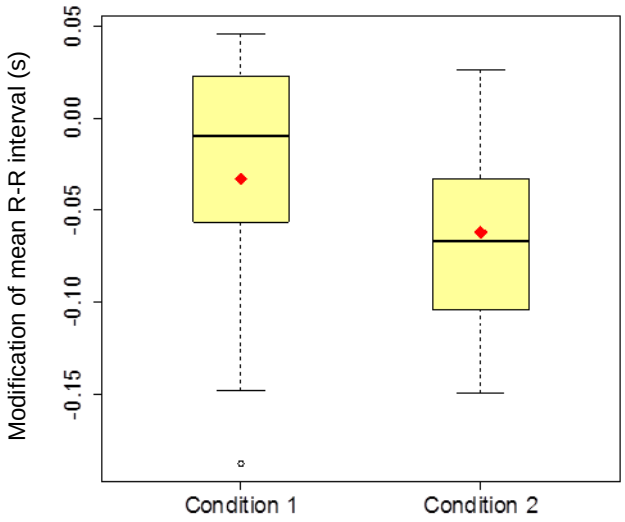


Figure 4. Differences between the mean R-R interval during the fire decision windows and the mean R-R interval during the minute that precedes the detection of the threat, by condition.

To conclude, there is no clear increase in heart rate during the decision making process in the first condition (C1) whereas this increase is significant with the late communication strategy (C2). Although this result is not confirmed by the direct comparison between both conditions, it is therefore likely that with a larger number of subjects, the difference between the two conditions should become significant.

For the electrodermal activity, the statistical analysis of measure 1 indicates that in both conditions an increase in skin conductance is observed (Paired t-test; C1: t = -5.920, df = 9, p-value = 0.000, Cohen's *d* = 2.135; C2: t = -4.300, df = 9, p-value = 0.002, Cohen's *d* = 1.372). Nevertheless, no significant difference appears between the two experimental conditions (Two sample t-test; t = -0.566, df = 18, p-value = 0.578, Cohen's *d* = 0.253). Thus, the detection of the new threat, and the opportunity to attack it, implies an activation of the sympathetic branch of the autonomic system from the operator. Moreover, this physiological response seems to be equivalent, whatever the communication strategy of the UCAV.

When measurement 2 is analysed, the skin response related to the drone request during the firing window is slightly different for the two experimental conditions, although the predefined statistical level is not reached (Two sample t-test; t = -1.554, df = 18, p-value = 0.138, Cohen's *d* = 0.695). It is likely that with some additional participants these differences become significant as the Fligner-Policello test (U* = 2.465, p-value = 0.015) considers that the difference is significant. The overall increase in skin

conductance can therefore be considered steeper in the second condition and more concentrated during the decision-making process.

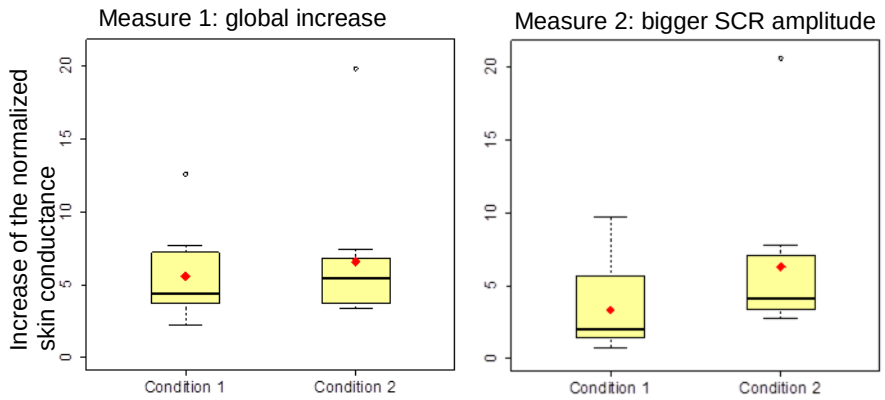


Figure 5. Modification, by measure and condition, of the normalized skin conductance related to the threat detection and fire decision-making process.

In conclusion, the physiological parameters indicate that the early communication strategy slightly reduces stress levels and workload, with a smaller increase in heart rate and an increase in skin conductance that spread out over time.

Testing hypothesis 2

The second hypothesis is that the early communication strategy should improve the operators' performance in making a firing decision. Two indicators are considered here, the validity of the response and the reaction time. In this experiment, a complete detection and identification of all elements in the scene, as well as an accurate application of the rules of engagement, should lead the operator to request a firing clearance from headquarters. Of the three possible actions (fire, request permission, abort), the first is a clear violation of the rules of engagement, the second is the expected one and the third is sub-optimal. Table 1 shows the results of the experiment.

Table 1. Summary of operators' decisions, by condition

	Condition 1	Condition 2
Fire	1	6
Request permission	8	4
Abort	1	0

A χ^2 test with this contingency table indicates that differences between the two conditions are just over the predefined threshold (X-squared = 5.905, df = 2, p-value = 0.052). It is very likely that with some additional participants these differences become significant. From an operational point of view, a majority of operators violated the rule of engagement in experimental condition 2, which is simply not acceptable.

Let us now look at the response time, calculated as the time between the request for a shot by the UCAV and the moment when the operator starts an action (fire, request permission, or abort). The t test (t = -1.409, df = 18, p-value = 0.176, Cohen's d = 0.630) indicates that the difference is not significant. From an operational point of view, the average response time is 4.5 seconds shorter in condition 1 (Fig. 6), which can be a real advantage in case of hostile enemy reaction.

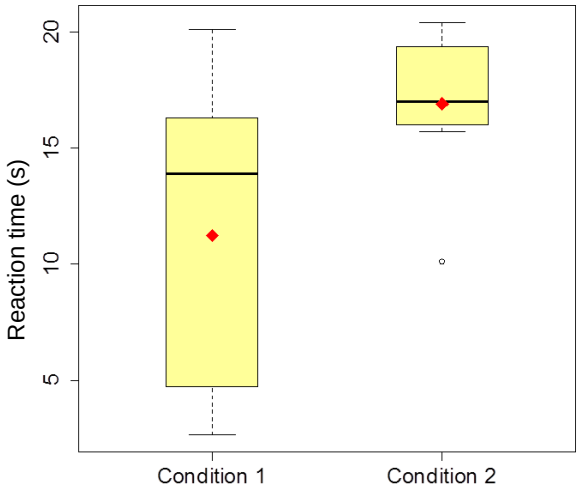


Figure 5. Reaction time by condition

Thus, these results broadly support the early communication strategy (C1) where better decisions are made in a shorter time. Nevertheless, other operators should be taken into consideration to confirm the robustness of these results.

Discussion

This study focuses on issues related to supervision and critical cooperative decision making involving an operator and a highly automated system. With the increasing level of automation and autonomy and the diffusion of more and more sophisticated objects in everyday life, it is of interest to better understand and assess their impact on human decisions. A plausible military scenario was chosen to outline a situation where the operator has to take, under strong time pressure, the responsibility of a critical decision that has been largely prepared by the partially autonomous system. Implemented in an immersive UCAV simulator, an ecological experiment was conducted to better understand the impact of the UCAV's communication strategy on the decision-making process and the overall monitoring of the automated system. This article only reports the results related to critical decision making.

Firstly, the physiological parameters recorded indicate that the decision-making process is always accompanied by an increase in electrodermal response and heart rate, reflecting an increase in stress level and workload. However, when the autonomous system warns the operator that a major decision has to be made soon, the increase in stress level (as captured by the electrodermal response) is spread over the

available time. Heart rate analysis indicates that the overall increase in stress and workload appears to be slightly lower in this condition.

Secondly, the communication strategy has a real impact on the final decision and far fewer wrong decisions are made when the operator knows a few minutes before that he will have to assess a situation and make a decision. With the early communication strategy, the operator has a better perception of the elements of the battlefield and makes more accurate use of the rules of engagement. In addition, the reaction time is shorter.

It is also found that the warning process modifies the operator's level of alertness and vigilance. Thus, although the time windows in which the operator can acquire the required information and construct his decision are exactly the same, the operator is more effective when he/she has been prepared to act. An interesting observation is also that, in the late communication strategy, the wrong decision was always to fire (rather than to abort the attack), as if the strong time pressure pushed the operator to follow the system's decision. Such a result needs to be studied with other experiments, but it is consistent with the notion of complacency towards automation that has already been studied in the aeronautical field. Finally, analyses of QUASA, NASA-TLX and debriefing data (not presented in this paper) indicate that there is no real difference in the overall mission. Users do not report that the late communication strategy is more uncomfortable and do not realise that their decisions were not in line with the rules of engagement.

Further studies are now needed to reinforce these results but also to determine how to reduce complacency towards automation. One perspective is to work on how the autonomous system can better 'explain' to the end user what is relevant to the decision.

References

Barnes, M.J., & Evans A.W. (2016). Soldier-Robot Teams in Future Battlefields: An Overview. In F. Jentsch, and M. Barnes (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 9-30), New York: Routledge.

Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., Dawson, M.E., & Filion, D.L. (2012). Publication Recommendations for Electrodermal Measurements. *Psychophysiology*, *49*, 1017–1034.

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

Eriksson, A., & Neville A.S. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. *Human Factors, 59*, 689–705.

Hancock, P.A., & Szalma, J.L. (2008). Stress and Performance. In P. Hancock and J. Szalma (Eds.), *Performance under Stress* (pp. 1-18), Aldershot: Ashgate.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology, 52*, 139-183.

Huang, G., & Pitts, B.J. (2022). Takeover Requests for Automated Driving: The Effects of Signal Direction, Lead Time, and Modality on Takeover Performance. *Accident Analysis & Prevention, 165*, 106534.

Mayer, M. (2015). The New Killer Drones: Understanding the Strategic Implications of next-Generation Unmanned Combat Aerial Vehicles. *International Affairs 91*, 765–780.

McGuinness, B. (2004). Quantitative Analysis of Situational Awareness (QUASA): Applying Signal Detection Theory to True/False Probes and Self-Ratings. In ICCRTS. Copenhagen, Dk.

Mirri, S., Prandi, C., & Salomoni, P. (2019). Human-Drone Interaction: State of the Art, Open Issues and Challenges. In *MAGESys'19 Proceedings* (pp. 43–48). Beijing, China: ACM.

Morgan, P.L., Alford, C., & Parkhurst G. (2016). *Handover Issues in Autonomous Driving: A Literature Review* (Project Report). Bristol, UK : University of the West of England.

Morgan, P.L., Alford, C., Williams, C., Parkhurst, G., & Pipe, T.. (2018). Manual Takeover and Handover of a Simulated Fully Autonomous Vehicle Within Urban and Extra-Urban Settings. In N. Stanton (Eds.) *Advances in Human Aspects of Transportation* (pp. 760–771). Cham: Springer.

Olsen, D.R., & Goodrich, M.A. (2003). Metrics for Evaluating Human-Robot Interactions. In PerMIS'03 Worksop Proceedings. Gaithersburg, MD, USA.