Make automation G.R.E.A.T (again)

Julien Donnot,[1] Daniel Hauret,[2] Vincent Tardan,[1] & Jérôme Ranc.[1]
[1]CEAM, Mont-de-Marsan, France
[2]Thales AVS, Mérignac, France
julien.donnot@intradef.gouv.fr

INTRODUCTION

The objective of the current work was to test a method designed for the evaluation of human trust behaviors in Artificial Intelligence Based System (AIBS). The French Air Force needs to develop the capability to evaluate usability of AIBS employed on board of a fighter aircraft. Trust in autonomy can be considered as a key factor for usability. As French Air Force has to be prepared for forthcoming conflicts, conception of future air combat systems implies to anticipate ergonomics issues, especially in terms of decision-making, workload and errors in fighter aircraft cockpits. In the field of prospective ergonomics (Robert & Brangier, 2009), Brangier and Robert (2014) point out the difficulty to represent future activity related to a system that does not yet exist. Considering that the main characteristic of future fighter aircrafts will be the obligation for pilots to collaborate with an AIBS (Lyons, Sycara, Lewis & Capiola, 2021), there is a real need to think up this future collaborating activity.

Lyons et al. (2018) warned about the specificity of trust in future autonomy, including AIBS, in the field of military aviation. Leading studies with real operators, with real tools and real consequences (R3 concept) appears as the most relevant. In the field of military aviation, real operators are fighter pilots, real tools are fighter aircraft (Rafale) and real consequences appear in a tactical environment. Too few studies reported knowledge about the French fighter pilot activity. Amalberti (1996) touched on some specific features of this activity, Guérin, Chauvin, Leroy, and Coppin (2013) adapted a Hierarchical Task Analysis method to one air operation and Hauret (2010) was the first to be interested in pilot collaboration with an artificial agent.

To define what would be the collaborating activity in a future fighter cockpit, ergonomists need to assure usability of human machine interfaces. Bastien and Scapin (1995) described a set of criteria designed for conception guidance. These guidelines were thought to conceive human-computer interfaces. Given that functions performed by AIBS are and will be more complex and sometimes innovative, AIBS conception guidelines deserve to be considered. In the current study, authors focused on trust as a critical factor for usability of AIBS on board a fighter aircraft. Then, the objective of the study was to produce conception guidelines to increase usability by building pilot's trust in AIBS.

Trust is a complex concept depending on individual, organizational and cultural context (Lee & See, 2004) but we choose to focus on its calibration in the current study by considering the lack and the excess of trust in a specific AIBS. A large number of methods and metrics can drive analysis of trust levels (Hoff & Bashir, 2015). In order to assess usability in relation with trust, pilots' behaviors prevailed over pilots' feelings. Therefore, we develop a method immersing operational fighter pilots in a simulated combat air mission with an operating AIBS. Experimental objectives were 1) to identify causes of observed trust levels leading to understand pilots uses of the AIBS and 2) to formulate ergonomics principles justified by trust issues.

METHODOLOGY

Participants

Four military pilots were tested. All the pilots were experimented on Mirage 2000 and familiar with aircraft simulator.

Apparatus

Participants were asked to perform a flight as an operational mission on a Mirage 2000 simulator. Mission was built and played on DCS world®. Functions of the AIBS PathOptim® were integrated into a Tacview® interface.

The mission demands each pilot to cross a hostile territory to bomb a target on time and to come back safe. Pilots had to respect several restrictions like a maximal height of 10kft, no detection by enemy radar and no engagement by enemy air defense systems. Flying over 10kft for the first time made an enemy fighter aircraft took off for interception and raised mission difficulty.

The AIBS

PathOptim is a 3D track solver based on the Genetic Fuzzy Trees method, which gives to the pilot the choice of three types of track to reach the target as fast as possible. By integrating characteristics of enemy air defense systems, each track is calculated for a fixed minimal height without overflying 10kft at a fixed speed. Green tracks avoid as much as possible to enter in SAM[1] ring, red tracks are the fastest tracks even if the pilot must enter in one or several SAM rings and amber tracks are a compromise between survival and fastness (Figure 1).

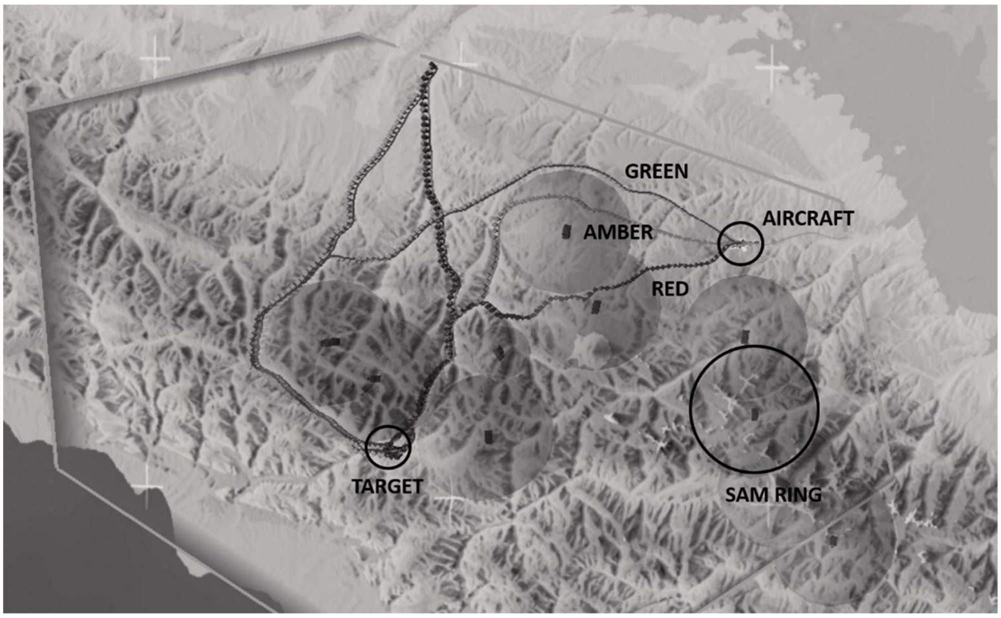

Figure 1: PathOptim in the tactical display

Scenario

During the mission, some new air defense threats appear if the pilot enters in the triggering area. Events could be SAM appearance, SAM disappearance and a new threat not considered by PathOptim (soldiers with Rocket Launchers - RPG). Pilots were informed of PathOptim limitations. Events were built to force pilots to use PathOptim. As a control condition and the only exception, the threat by soldiers with RPG, was created to confirm that pilots were aware of the uselessness of PathOptim is this specific situation.

Pilots must react as following: Refresh PathOptim, choose a track and follow the track. Trust coding results from the combination of the possible behaviors produced by pilots (Figure 2). Pilots could use PathOptim at any time even if no tactical change pops up.

---

[1] Surface Air Missile

The experiment setup was designed to deduce trust, undertrust and overtrust from PathOptim uses by pilots.

Data analysis

Each planned events and each supplementary use of PathOptim were analysed. Analysis consisted in replaying the mission. Tactical situation, aircraft spatial localization, pilots/cockpit interactions, pilots/PathOptim interactions, eye-tracking and radio communications were analyzed to understand pilots' behaviors. The workload was monitored according to the dual-task paradigm. Pilots had to add a triplet of digits and give the total. Data were analyzed in terms of accuracy and delay in seconds.

Trust was observed when pilots used PathOptim and that PathOptim was helpful (Figure 2). With understrust, the pilot did not use it whereas PathOptim was helpful. In undertrust we distinguished defiance and distrust. With defiance, the pilot refreshed PathOptim but did not follow a proposed path and in distrust, the pilot did not refresh PathOptim nor followed a virtual path. In overtrust, the pilot did not refresh PathOptim although he needed to do it. Another overtrust situation was when the pilot refreshed PathOptim whereas it would not be useful (i.e., soldiers with RPG not detected by PathOptim).

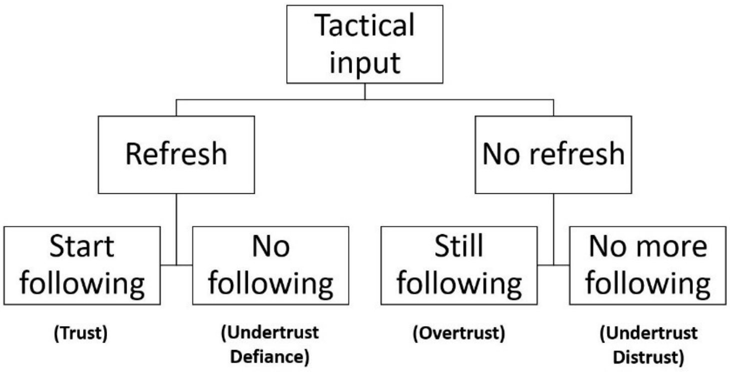

Figure 2: Hierarchical action tree and corresponding trust levels.

FINDINGS AND DISCUSSION

General results

All the pilots achieved the mission successfully by bombing the target and without being killed by enemy air defense.

Trust calibration

Pilots followed a path proposed by PathOptim between 62% and 73% of the flying time. The number of PathOptim utilizations varied according to the pilots and produced from six to thirteen events. Trust represented from 33% to 82% of trust levels against undertrust and overtrust. Trust tends to increase with number of events and therefore with uses.

Regarding the workload, high delays can be explained by enemy's aircraft monitoring, integration of new events in the situational awareness and path following.

Qualitative analysis led to identify effects and causes for each event generating undertrust or overtrust. Thus, ergonomists produced requirements for PathOptim development. Based on these requirements, ergonomic principles have been formulated.

G.R.E.A.T Principles

G for Guidance

The pilot needs the necessary information to follow the tracks proposed by PathOptim, so AIBS must carry out tasks helping the proposal execution (ex: diving tight curve) to reduce flying errors from the pilot.

R for Recommendation and Realism

Recommendation: The pilot needs information on the currently best use of the system, so AIBS must be able to detect inappropriate use by the pilot (ex: no green path because amber path was forced) to make use of the full system capabilities.

Realism: The AIBS has to treat environmental factors responsible for an effect on security or performance to gain trust with a well-fitting proposal (ex: flying in the valley instead of above hills).

E for margin of Error

The pilot needs to be aware of the margin allowed for execution of the AIBS proposal (ex: the aircraft has to be less than 100m from the virtual path). Thus, AIBS must inform the pilot about the conditions of the validity proposal to reduce interpretation errors and gain trust for tactical decision-making and track following (ex: best path as long as the pilot is no more than 5 sec late = low margin of error).

A for Automation

AIBS must treat automatically environmental and tactical changes to relieve the pilot from considering changes.

T for Transition

AIBS proposals must tend to be univocal and understandable to reduce pilot's doubt and speed up decision-making.

Conclusion

G.R.E.A.T principles present the benefit of being justified by real use cases. These principles will probably evolve as long as ergonomists experiment new and various use cases.

REFERENCES

Amalberti, R. (1996). *La conduite des systèmes à risques*. Paris : PUF.

Bastien, J. M. C., & Scapin, D. L. (1995). Evaluating a user interface with ergonomics criteria. *International Journal of Human Computer Interaction*, *7*, 105-121.

Brangier, E., & Robert J.-M. (2014). L'ergonomie prospective : fondements et enjeux. *Le Travail Humain, 77*, 1-20. doi: 10.3917/th.771.0001

Guérin, C., Chauvin, C., Leroy, B., & Coppin, G. (2013). Analyse de la tâche d'un pilote de Rafale à l'aide d'une HTA étendue à la gestion des modes dégradés. Communication présentée à la 7ème conférence Epique, Bruxelles, Belgique. (p.21-26).

Hauret, D. (2010). *L'autorité partagée pour mieux décider à bord du Rafale.*, Communication présentée à la 12ème conférence internationale ERGO'IA 10, Bidart, France : ACM Press

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*, 407–434. https://doi.org/10.1177/0018720814547570

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46, 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lyons, J. B., Ho, N. T., Van Abel, A. L., Hoffmann, L. C., Fergueson, W. E., Sadler, G. G., et al. (2018). “Exploring trust barriers to future autonomy: a qualitative look,” in *Advances in Human Factors in Simulation and Modeling, Advances in Intelligent Systems and Computing 591*, ed. D. N. Cassenti (Cham: Springer).

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). *Human–Autonomy Teaming: Definitions, Debates, and Directions.* Front. Psychol. 12:589585. doi:10.3389/fpsyg.2021.589585

Robert, J-M., & Brangier, E., (2009). « What is prospective ergonomics? A reflection and position on the future of ergonomic”, in B.-T. Karsh (Ed.): *Ergonomics and Health Aspects,* LNCS 5624, pp. 162–169.