Integration of Policy Capturing for Online Support of Pilot Decisions in an Aircraft Simulation

Ramon Alaman, J.[1, a], Marois, A.[2], Salvan, L.[2], Lavoie-Hudon, L.[1], Lafond, D.[2], Yan, S.[3], Tremblay, S.[1]
[1]École de Psychologie, Université Laval, Québec City, QC, Canada
[2]Thales Research and Technology, Québec, Canada
[3] Thales, Toulouse, France
[a]jonay.ramon-alaman.1@ulaval.ca

INTRODUCTION

In the coming years, the aviation industry is expected to face several challenges, such as pilot shortages due to traffic increases or the need to reduce operational costs. This context points towards the development of Single Pilot Operations (SPOs; Myers & Starr, 2021). Such situation encourages research on systems capable of supporting pilots in their tasks because, otherwise, the pilot should assume alone the tasks carried out by two people (cf. Mosier et al., 2001). Even if SPOs are finally not adopted, the research and scientific discussion in this field would still be beneficial (Neis, Klingauf & Schiefele, 2018) and the introduction of SPO-supporting systems in traditional cockpits should translate into safer operation of aircraft. For instance, to develop procedures and technologies to assist the remaining pilot in case of incapacitation.

One possible solution would be automation, which could reduce the workload of the crew (or the pilot in SPOs; Tokadlı, Dorneich, & Matessa, 2021). However, the reduction of situation awareness due to automation is a well-documented problem in aviation (Sikora et al., 2020). This could be particularly relevant when the activities carried out by automated systems are not clearly notified to the crew, isolating them from the operation of the aircraft. Such a situation may lead to a divergence between the actual aircraft state and the pilot mental model of the aircraft (Silva & Hansman, 2015). Pilots should be given the necessary information to be able to properly calibrate their confidence on aircraft systems (Okamura & Yamada, 2020). Also, the automation bias, defined as the overreliance in automation and the replacement of a standard decision-making process by "the use of automated cues as a heuristic" (Mosier et al., 2017, p. 47), should be considered. Thus, automation needs to maintain the pilot in an active decision-making role with a sufficient level of readiness to react properly to any situation that may arise (Bilimoria et al., 2014).

The selected use case for this paper is a situation that illustrates this need for assistance in taking the proper action during the last segment of the flight. At an altitude of 1,000ft, the pilots need to evaluate the stability of the approach and decide whether to land the aircraft or perform a go-around, i.e., abort the landing. A landing approach is considered unstable when there are significant trajectory or speed deviations. This phenomenon was a factor in approximately 10% of all airplane accidents that occurred between 2011 and 2016 (IATA, 2016).

To cope with such critical situations, one possible solution is to generate an expert decision-making model able to support the pilots. This can be done through policy capturing, which uses statistical or machine learning methods to model the human decision-making process and generate a decision from a set of predictors (Lafond et al., 2017). Thales Research and Technology Canada developed Cognitive Shadow, a prototype system that models the decision-making patterns of the user, updates this model continuously, and notifies the user of any discrepancy between the model and the user's current decision (Lafond et al., 2020). The addition of Cognitive Shadow suggestions can generate a joint human-machine cognitive system that is more precise than each of its components (Lafond et al., 2017). The resulting system can learn continuously and automatically, which is a key feature of cognitive systems (Christensen et al., 2010). Cognitive Shadow has been previously applied to simulation of maritime decision support (e.g., Labonté, et al., 2020) and to gaming with "opponent modeling" (Lavoie-Hudon et al., 2021).

The current paper presents ongoing work aimed at developing a policy capturing tool to support pilots' decisions in high-fidelity, dynamic situations. Firstly, we report our findings from a feasibility study involving modeling critical decisions during an unstable approach simulation. Secondly, we discuss key human factors and cognitive engineering issues to be addressed to de-risk the application of the Cognitive Shadow solution in the avionics domain.

PILOT EXPERIMENT

Method and Procedure

Four expert (mean experience = 18.8 years) and fit-to-fly pilots participated in the experiment. For 3h each, they faced as many unstable approach simulations as possible. The scenarios were selected from a set of 24 unstable approaches of which half were prone to a go-around decision and half to a landing decision. Terrain and meteorological conditions i.e., wind direction and velocity as well as presence or absence of clouds, precipitations, and wind shear changed across the different simulations. These scenarios and conditions were distributed quasi-randomly through simulations to ensure that participants were exposed to different situations. The simulations were run in X-Plane 11. The aircraft was a typical single aisle model. As shown in Figure 1, two screens were used to display the simulation. The one on the right displayed the out of the window view and the one on the left displayed a set of flight instruments: Electronic Flight Instrument System and its control panel, Flight Control Unit, Multi-function Control Display Unit. Data from the simulation was sent via UDP protocol and collected by a real-time processing nexus app (Sensor Hub; Gagnon et al., 2014). It was then sent through web socket to a tailored app for preprocessing and then, sent to the Cognitive Shadow web app via API REST to be recorded as features to allow capturing expert judgment policies.

Figure 1: Experimental setup.

Following the explanation of Cognitive Shadow and the experimental procedure, the experimenter started the simulation. Pilots were asked to perform an approach that would lead to the selected result, i.e., a go-around or a landing. At an altitude of 1,000ft, the experimenter would stop the simulation and asked the pilot for their decision, i.e., landing or going around, while considering the current state of the situation. If the pilot decided to go-around, a new simulation was started. Otherwise, the simulation continued until the aircraft was at 500ft altitude where the simulation was again paused, and the experimenter asked the pilot about their decision on whether to go-around or to land.

Results and Discussion

The number of scenarios faced by each pilot ranged between 5 and 30 decisions for a total of 65, of which 29 were landing continuations and 36 were go-arounds. Data from two decisions could not be retrieved. We then established a group model comprised of all the decisions of the four pilots, containing 63 decisions (27 go-around and 36 continue). Table 1 depicts the results of each of the seven Cognitive Shadow models. Given the low number of decisions, the accuracies of the models were computed by a leave-one-out cross-validation approach. The support vector classifier was the best performing, reaching an accuracy of 99.8%. It should be borne in mind that due to the low number of decisions we could not validate the models with an unseen test dataset (ie., although the models learned only from the training dataset, the hold-out validation dataset did inform hyperparameter optimization). Thus, there is a risk of model overfitting and replication with a larger dataset in future studies is necessary. However, similar accuracy levels (i.e., above 95%) were reached by the Cognitive Shadow in a high-fidelity simulation in the maritime surveillance domain with such small sample sizes (Chatelais et al., 2020). Despite this limitation, the present results demonstrated the feasibility of collecting live data from a high-fidelity simulator to feed a policy capturing system for real-time cognitive modeling and decision support.

Model	Global accuracy	Decision	Precision
Naïve Bayes	58.7	Continue	58.0
		Go-around	100
Decision tree	95.8	Continue	94.6
		Go-around	97.6
K-nearest neighbor	97.4	Continue	97.4
		Go-around	96.4
Support vector classifier	99.8	Continue	99.7
		Go-around	100
Logistic regression	67.3	Continue	65.0
		Go-around	77.9
Neural network	88.2	Continue	84.8
		Go-around	94.6
Random forest	91.3	Continue	88.5
		Go-around	96.0

Table 1: Global accuracy and decision-dependent precision (in %) of the seven models generated by the Cognitive Shadow to predict pilots' decisions in the X-Plane 11 simulation.

SCIENTIFIC DE-RISKING FOR AVIATION APPLICATIONS

Future work will focus on the effects and challenges of introducing in the cockpit support systems such as Cognitive Shadow from a human factors' perspective. Particularly relevant are the calibration of confidence (McGuirl & Sarter, 2006; Okamura & Yamada, 2020), the automation bias (Mosier et al., 2017), and several human factors-related issues such as the maintenance of situation awareness and the optimal format of suggestions to ensure adequacy to the situation and pilots' understanding (Degas et al., 2022).

Confidence in decision support systems depends not only on their design, but also on the users and on the environment in which they navigate (Hoff & Bashir, 2015). Thus, the means to build confidence in an automated system need to adapt to the specific characteristics of the studied situation. These means include ensuring the user's comprehension of the system and its suggestions or providing ongoing confidence feedback (Hoff & Bashir, 2015). However, each of these methods has its own weaknesses. Using the system and observing the absence of error may lead to overconfidence (McGuirl & Sarter, 2006). The use of confidence metrics has been shown to improve decision-making of expert pilots (McGuirl & Sarter, 2006) yet was found to have no impact on novices (Lafond et al. 2020). Thus, further research is needed to better understand the factors and dynamics leading to appropriate levels of trust.

Cognitive Shadow is proposed as a warning system, intervening only when a pilot deviates from the expected decision pattern as the intent is to keep the pilot in charge. This functioning has shown its benefits compared to a continuous suggestion mode (Labonté et al., 2020). Model output justifications based on SHapley Additive exPlanations (SHAP) values have been added to the Cognitive Shadow's feedback capability. These values are a measure of each factor's importance on the system output (Lundberg & Lee, 2017) aiming at helping the pilot understand the situation parameters deemed critical by the model and take a more informed decision.

It is also of relevant in the development of cognitive assistants to study their potential use in pilot training. The use of systems based on models of expert behavior capable of giving online feedback to pilots in training is certainly an option worth exploring. Policy capturing systems may help by identifying weaknesses in the trainees' decision-making, but also by providing timely online suggestions and justifications on what experts would have done. The use of Cognitive Shadow in training situations will need a specific evaluation of the previously analyzed design factors associated with it.

CONCLUSION

Modelling the decision-making process of aircraft pilots allows developing tools capable of supporting and training them. In this paper, we presented the results of a feasibility study showing the ability of such a system to model the decision-making process of a small group of pilots in a critical situation, i.e., an unstable approach. Despite the minimal number of decisions used for training, preliminary results showed great promise in terms of predictive accuracy and demonstrated the technical feasibility of online policy capturing in a high-fidelity dynamic environment. Different human factors aspects need to be addressed to integrate such a tool to operational situations. How this system should be integrated into the cockpit and how it could be used to train future pilots are key questions to be investigated in future studies.

REFERENCES

Bilimoria, K. D., Johnson, W. W., & Schutte, P. C. (2014). Conceptual framework for single pilot operations. In *Proceedings of the international conference on human-computer interaction in aerospace* (pp. 1-8).

Chatelais, B., Lafond, D., Hains, A., & Gagné, C. (2020). Improving policy-capturing with active learning for real-time decision support. In *Proceedings of the International Conference on Intelligent Human Systems Integration, 3,* 177–182.

Christensen, H. I., Sloman, A., Kruijff, G. J., & Wyatt, J. L. (2010). Cognitive systems introduction. In *Cognitive Systems* (pp. 3-48). Springer, Berlin, Heidelberg.

Degas, A., Islam, M.R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M.U., Begum, S., Rahman, M.A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G.,

Babiloni, F., & Aricó, P. (2022). A Survey on Artificial Intelligence (AI) and eXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory. *Applied Sciences*, 12(3), 1295.
Gagnon, J. F., Lafond, D., Rivest, M., Couderc, F., & Tremblay, S. (2014, May). Sensor-Hub: A real-time data integration and processing nexus for adaptive C2 systems. In *Proceedings of the sixth international conference on adaptive and self-adaptive systems and applications* (pp. 63-67).
Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3), 407-434.
International Air Transportation Association (IATA). (2016). Unstable Approaches: Risk Mitigation Policies, Procedures & Best Practices.
Labonté, K., Lafond, D., Hunter, A., Neyedli, H. F., & Tremblay, S. (2020, December). Comparing two decision support modes using the Cognitive Shadow online policy-capturing system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 1125-1129). Sage CA: Los Angeles, CA: SAGE Publications.
Lafond, D., Labonté, K., Hunter, A., Neyedli, H. F., & Tremblay, S. (2019, August). Judgment analysis for real-time decision support using the Cognitive Shadow policy-capturing system. In *International Conference on Human Interaction and Emerging Technologies* (pp. 78-83). Springer, Cham.
Lafond, D., Roberge-Vallières, B., Vachon, F., & Tremblay, S. (2017). Judgment analysis in a dynamic multitask environment: Capturing nonlinear policies using decision trees. *Journal of Cognitive Engineering and Decision Making*, 11(2), 122-135.
Lavoie-Hudon, L., Lafond, D., Labonté, K., & Tremblay, S. (2021, October). Cognitive Shadowing for Learning Opponents in a Strategy Game Experiment: Using Machine Learning to Counter Players' Strategies. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play* (pp. 308-313).
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
Myers, P. L., & Starr, A. W. (2021). Single Pilot Operations IN Commercial Cockpits: Background, Challenges, and Options. *Journal of Intelligent & Robotic Systems*, 102(1), 1-15.
McGuirl, J. M., & Sarter, N. B. (2006). Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information. *Human Factors,* 48(4), 656–665. https://doi.org/10.1518/001872006779166334
Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (2017). Automation bias: Decision making and performance in high-tech cockpits. In *Decision Making in Aviation* (pp. 271-288). Routledge.
Neis, S. M., Klingauf, U., & Schiefele, J. (2018, September). Classification and review of conceptual frameworks for commercial single pilot operations. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)* (pp. 1-8). IEEE.
Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *Plos one*, 15(2), e0229132.
Sikora, I., Hari, B. L., & Hanusch, M. (2020). Human factors approaches and models in LOC-I accident analysis and prevention: flight crew resource management techniques as a risk mitigation tool. *International Journal of Safety and Security Engineering*, 10(3), 301-310.
Silva, S. S., & Hansman, R. J. (2015). Divergence between flight crew mental model and aircraft system state in auto-throttle mode confusion accident and incident cases. *Journal of cognitive engineering and decision making*, 9(4), 312-328.
Tokadlı, G., Dorneich, M. C., & Matessa, M. (2021). Evaluation of Playbook Delegation Approach in Human-Autonomy Teaming for Single Pilot Operations. *International Journal of Human–Computer Interaction*, 37(7), 703-716.