



Experimental Evaluation of a Scalable Mixed-Initiative Planning Associate for Future Military Helicopter Missions

Fabian Schmitt^(✉) and Axel Schulte

Institute of Flight Systems, University Bundeswehr Munich,
Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany
{fabian.schmitt, axel.schulte}@unibw.de

Abstract. This article describes a scalable mixed-initiative planning concept, in which a human pilot is assisted during mission (re-)planning by an artificial planning agent. The agent serves as an additional team member and enables rapid planning and re-planning of multiple vehicles. For this purpose, the agent adapts its extent of assistance based on the necessity of the given situation. The concept was implemented for the use case of manned-unmanned teaming in future military helicopter missions. Thereby, the mixed-initiative agent was implemented with three different levels of automation. The article focuses on the experimental evaluation with German military helicopter pilots. Results show the advantages of the scalable mixed-initiative concept especially in time critical and high workload situations.

Keywords: AI systems · Associate systems · Mixed-initiative Problem solving

1 Introduction

The mission planning, re-planning and management of multiple vehicles in dynamic scenarios by a single pilot is a highly topical field of research. A major issue within this research field is the risk for excessive operator *mental workload (MWL)* and loss of operator *situation awareness (SA)*. Systems for partly or fully automated planning and scheduling can solve complex multi-vehicle problems in reasonable time. However, these systems are often not directly suitable for use in user-oriented incremental and collaborative planning [1]. Instead, such systems are developed to perform their dedicated function. This might result in a number of fundamental human factors related problems. A loss of *situation awareness (SA)* and *plan awareness (PA)* is likely to occur in critical situations, because the operator is not integrated into the planning process anymore. Furthermore, complacency and over-reliance may occur, because operators may overly trust those planning systems.

In order to counteract such issues, we propose a *mixed-initiative (MI)* planning approach. We define *mixed-initiative* as a cooperative approach between a human operator and a cognitive agent to solve a common planning problem. Thereby, both the

human operator and the agent can take initiative over the planning process and direct the process into a certain direction.

In this article, we present a *scalable mixed-initiative* planning agent. This agent is able to adapt the extent of its intervention to the situation. The emphasis of this article is on its experimental evaluation. In the following two sections, we will describe firstly our application and secondly the concept of this agent. The subsequent section describes an experimental campaign, which was conducted recently with German military helicopter pilots to evaluate the agent in complex mission scenarios. The article concludes with results of that research campaign.

This work builds on research originated in [1]. Multiple other *mixed-initiative* approaches were already developed in [2–4] for various domains. However consequent experimental evaluations are rare.

2 Application

In our research, we look at *Manned-Unmanned Teaming (MUM-T)* in the field of military helicopter missions. Here, a manned two-seated helicopter is teamed with three small Unmanned Aerial Vehicles (UAVs). The UAVs serve as detached sensor platforms. They are designed to conduct route- and area reconnaissance and are able to detect safe landing points in hostile environments. Operational benefits include an increased sensor range of the manned platform, a better understanding of the tactical situation, increased lethality for the human pilots and eyes on target capabilities. The unmanned systems are controlled directly out of the cockpit by one of the pilots to reach a high level of interoperability. Thereby, pilots can access new reconnaissance information instantaneously and react rapidly if required. In our concept, the pilot in command (PIC, non-flying) serves as battlefield manager. He is responsible for the tactical planning and re-planning of the manned/unmanned team during the mission. The pilot's planning task consists of helicopter route planning (primary route and alternative route), identification and assignment of reconnaissance tasks to the UAVs and temporal coordination between different vehicles. Therefore, he works cooperatively with our *mixed-initiative* planning agent. The agent proposes future helicopter routes and UAV tasks, identifies threat conflicts in the current plan and offers suitable solutions. Furthermore, it helps optimizing a given plan. The communication between pilot and agent is dialog based. When the agent comes up with a recommendation, the pilot can either accept, reject or ignore the recommendation.

3 Concept

3.1 Work System

We will use a formal graphical description method, designed for *Human-Autonomy Teaming (HAT)* [5], to present the *mixed-initiative* approach. This method was developed our institute in order to structure top-level *HAT* designs and is based on the work process. Based on a given work objective and the corresponding work

environment, a certain work output shall be generated. The work system differentiates between the worker on the left-hand side and the tools on the right-hand side. The worker has knowledge of the overall work objective and tries to reach that objective by own initiative. The human worker is furthermore authorized to change this work objective by own initiative. To achieve that work objective, the worker uses given tools which are shown on the right-hand side of the work system. The tools are subordinates, i.e. hierarchically degraded with respect to the worker. More general information about the work process and work system notation can be found in [5]. The conceptual design of our *mixed-initiative* work system is presented in Fig. 1. On the worker (left-hand) side is the pilot in command, who is responsible for airborne mission (re-)planning. Three dislocated UAVs (right-hand side) serve as tools and can be used to fulfill the mission objective. A cognitive agent on-board each UAV provides a delegation interface to the pilot to enable task-based guidance and controls the actual UAVs in supervisory control. The human pilot uses degraded planning tools to (re-)plan or modify the mission plan whenever required.

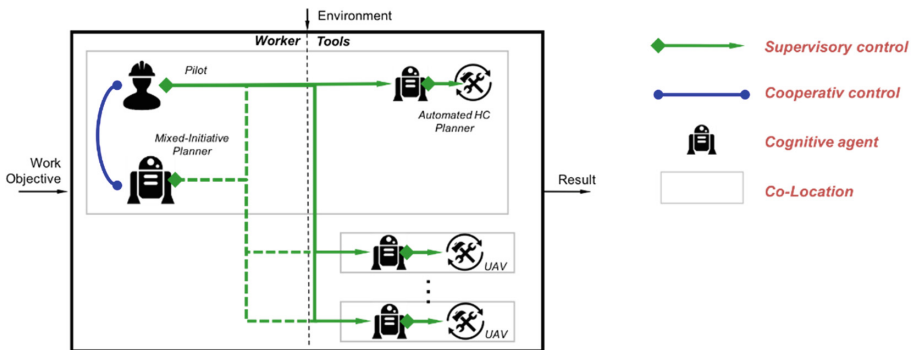


Fig. 1. Work system notation for the *mixed-initiative* approach

Additionally, we added a cognitive agent next to the pilot on the worker side, which represents the *mixed-initiative* agent. As a consequence of the placement on the worker side, the agent has to know the mission objective and can intervene in the planning process by own initiative. Thereby, initially the pilot provides information about a mission goal to the agent. Based on that, the agent has access to the described tools (i.e., UAVs and helicopter route planner) in the same way as the human pilot.

3.2 Agent Behavior

The role of the planning agent is to assist the pilot in the mission planning process. It shall ensure, that a satisficing and valid mission plan is available at all times throughout a mission. As long as the pilot is doing a good job at moderate workload, the agent shall remain passive as there is no need for any assistance. However, when the pilot needs any support, it shall identify future mission tasks as well as conflicts in the current plan and proposes suitable solutions to the pilot. It is important to point out, that

the pilot cannot assign tasks to the planning agent. Rather, the planning agent decides on its own, when to intervene.

We formulated the following behavior rules for the planning agent [6]:

- intervene as little as possible,
- intervene as late as possible (to enable the pilot to solve the problem on his own, if possible),
- leave as much work to the pilot as possible to maintain pilot's SA (i.e. adapt the extent of assistance/intervention to the pilot's workload),
- intervene incremental, rather than complex, whenever possible (a complex proposal might be hard to understand, *plan awareness* might diminish).

When designing the agent's behavior, we also oriented on Herbert Simon's *Satisficing Principle* [7] from organization psychology which describes the behavior of human decision makers. Simon emphasized especially two points:

1. The human in general does not try to find an optimal solution to a given problem. Instead, he stops working on the problem as soon as the solution is sufficient for him.
2. Sufficient quality depends on the decision maker's personal level of aspiration. The level of aspiration is thereby determined by the criticality of the given situation.

Based on Simon's principle our agent shall weigh if an interaction is really required in a given situation. Also, in low workload situations, the pilot might have an increased level of arousal. We formulated the following additional behavior rules:

- aspire after a sufficient rather than an optimal plan (the pilot does not aspire after an optimal plan), and
- adapt the aspiration level to the current tactical situation, and the pilot's mental workload.

3.3 Scalable Mixed-Initiative

To be able to adapt the extent of intervention to the pilot's *mental workload*, we defined three *Levels of Automation (LOAs)*. Depending on the given situation, one of the levels might be most suitable. The three levels of automation are:

1. *Silent mode*: In this mode the agent is designed to leave all work to the pilot. It corrects only errors in the plan (low workload mode, only available on initial mission planning).
2. *Incremental planning mode*: 1. + the agent takes initiative on the expansion of the plan using single/incremental steps of interventions. (medium workload mode, basic mode during flight) The agent shall reduce the share of work of the pilot during the planning process.
3. *Extensive planning mode*: 2. + the agent intervenes with complex/extensive interactions comprising multiple planning steps at a single time. (high workload mode, only in critical situations with high task load) In this mode, the agent is able to propose task sequences for multiple aircrafts at the same time to achieve rapid plan progress.

We defined a default level respectively for initial planning (*silent mode*), re-planning due to changes of tactical situation and throughout the flight (*incremental planning mode*), and re-planning due to changes of the mission objective (*extensive planning mode*). If there are no further information available about the pilot's *mental workload*, one of these levels shall be chosen automatically depending on the given situation. However, ideally, in *scalable mixed-initiative* the extent of agent interventions (i.e. the level of automation) shall be adapted automatically to the pilot's current workload.

While *MWL* can, in general, be reduced by increasing the extent of an intervention, pilot's *situation-* and *plan awareness (PA)* might probably diminish. The probability for lacks in *plan awareness* increases, because the pilot's part in the cooperative planning process decreases. Balancing *PA* and *MWL* is therefore one of the crucial factors in *scalable mixed-initiative*.

3.4 Distribution of Tasks Between Pilot and Agent

In order to balance workload, task responsibilities can be shared between pilot and agent by default. In our application, the pilot shall be responsible for the planning primary helicopter routes and the task identification and assignment to UAVs. In the contrary, the planning agent shall be responsible for planning of alternative helicopter routes as well as scheduling (permanently substituting assistance). However, if the pilot is not satisfied with the agent's solution, he can modify it.

4 Functional Architecture and Implementation

This chapter briefly describes the implementation of the *mixed-initiative* agent in the helicopter mission simulator at the Institute of Flight Systems.

4.1 Mixed-Initiative Agent

Four key capabilities are required to achieve the desired supporting agent behavior. These are described in the following:

Capability for Reasoning and Planning in the Given Application Domain. This capability is used to correct flaws in the current plan or to determine future actions, which are required in order to reach the mission objective. Based on this capability, the agent can compare planning solutions with each other and thus help optimizing the mission plan.

We modelled our planning domain using a-priori knowledge. We used PDDL as modelling language, which is an action-centered language, which is widely used to solve planning problems [8]. Core of PDDL are actions with pre- and post-conditions that describe the applicability and the effects of actions. In our application, these actions

comprise helicopter and UAV specific actions. We use a PDDL planner which works based on our mission domain and a problem file. Additionally, we use a CPLEX planner for rapid task assignment, optimization and scheduling [9].

Capabilities for Activity Determination. Knowledge about the pilot's planning activities is helpful to determine if the pilot is already working on the mission plan or not. If he is already working on an existing flaw in the plan, the agent might not need to intervene anymore. Honecker et al. developed an evidential reasoning approach derived from Dempster-Shafer theory, which is used to infer the current pilot's actions [10].

Capabilities to Estimate Pilot's Workload. Knowledge about the pilot's current workload is used to override the default *Level of Automation* for a given situation in order to adapt better to the pilot's current mental state. Brand and Schulte [11] presented a method and implementation of a workload-adaptive associate system.

Capabilities for Intervention. Finally, the agent needs the ability to interact with the pilot based on the usage of all aforementioned capabilities. Therefore, it uses rules to prioritize planning activities (e.g. the detection of a landing point is more important than the reconnaissance of an alternative helicopter route). Furthermore, the pilot's activities and his workload must be taken into account. Based on that the *mixed-initiative* agent has to infer a course of actions to generate a mission plan in cooperation with the pilot. We used PDDL to model possible interactions and prioritizations as actions with pre- and post-conditions. Costs were assigned to each action to model the *Satisficing* principle.

More details of the implementation can be found in [6, 12].

4.2 User Interface

We developed an HMI that satisfies requirements for mission planning and communication between pilot and agent. The HMI has the following three components:

- a mission interface,
- a tactical map display, and
- a dialog interface.

The mission interface is used by the pilot to specify the mission objective. The tactical map display is used by the pilot to sketch the mission plan and command tasks to the UAVs. Therefore, the pilot uses an object-oriented context menu. The map display is also used by the agent to visualize plan proposals and alternatives (Fig. 2). The dialog interface is used primarily by the agent to communicate with the pilot. The agent uses pushes text messages to propose new tasks or solutions (Fig. 3). Each message contains the particular problem identified by the agent. Additionally, most messages contain a proposed solution for that problem and an option for the pilot to either accept or reject the proposed solution. The solution is also visualized graphically on the map display in magenta.

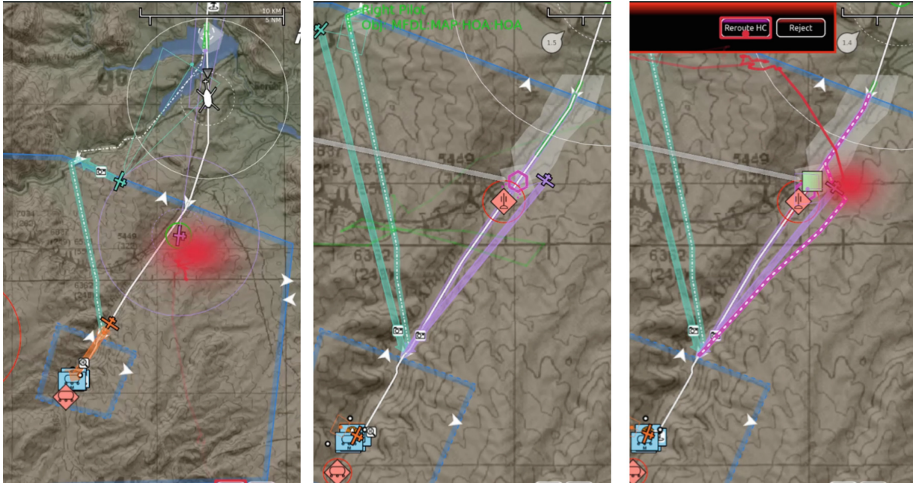


Fig. 2. Initial planning with preplanned alternative (left), pop-up threat on primary route (center) and re-planning due to threat (right). The red spots represent the pilot's gaze position. (Color figure online)

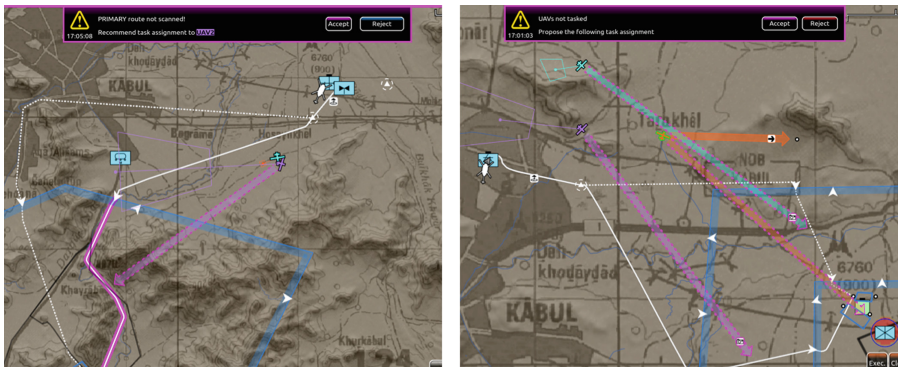


Fig. 3. Incremental task proposal (left) and extensive task proposal (right) of the agent.

5 Experimental Evaluation

5.1 Experimental Setup

A human-in-the-loop experiment was developed in order to evaluate our MUM-T configuration with crews of military pilots. Within this context, the cooperation between the pilot in command and the planning agent for the previously described three levels of automation was also evaluated. The experiment was conducted in the helicopter mission simulator at the Institute of Flight Systems (Fig. 4). In our experimental setup, the pilot non-flying was responsible for mission planning and re-planning of the manned helicopter and additionally 3 three unmanned systems. We exposed all crews

to multiple changes of the tactical situation as well as changes of the mission objective inflight. Thereby, the pilots were assisted by our *mixed-initiative* agent. During the missions, we observed pilot behavior using gaze tracking and manual system interactions). After each mission we gathered subjective ratings using prepared questionnaires.



Fig. 4. Helicopter mission simulator at the IFS

5.2 Participants

Eight military helicopter pilots from the German armed forces participated in the study. The subjects included 1 current and 7 former military pilots from 31 to 59 years of age ($M = 50.4$, $SD = 9.2$). Flight hours varied between 535 and 6850 ($M = 3933$, $SD = 1807$). The pilots were subdivided into four crews of two pilots each. Thereby, one of the pilots served as pilot in command (planning & mission management) while the other one served as pilot flying (responsible for control of the actual manned helicopter). PIC flight hours ranged between 2930 and 4920 h ($M = 3937$, $SD = 804$).

5.3 Scenario

We conducted five military MUM-T helicopter transport missions within three days. Each mission consisted of an initial planning phase on ground, an ingress phase into hostile territory, detection of pop-up threats, at least one landing in hostile territory, a change of the mission objective inflight, and an egress phase. Therefore, each mission contained at least a single use case for each level of automation:

- use case *silent mode*: initial mission planning on ground,
- use case *incremental planning mode*: change of tactical situation (pop-up threat on helicopter route), and
- use case *extensive planning mode*: change of mission objective.

The sequence of missions was kept equal across all crews. The designed scenarios were rated afterwards by each pilot using Likert scales (ranging from 1/negative to 7/positive). Realism of the scenarios was rated very well ($M = 5.6$, $SD = 0.7$). Mission sequence of actions was also rated positive ($M = 5.3$, $SD = 0.83$). Likewise, the simulation was rated rather positive ($M = 5.0$, $SD = 1.0$). Mission duration varied between 31 min and 71 min ($M = 45.77$ min $SD = 10.25$ min). Prior to the first mission, pilots were trained on the simulator for 2 full days. Thereby, the training consisted of a tutorial and 4 training missions with increasing complexity. Learning objectives were tested after completion of the tutorial for each pilot.

6 Results

We analyzed three different situations in each mission, which matched to our use cases (initial planning – silent mode, re-planning due to change of tactical situation, re-planning due to change of mission objective). In the following, we present results of objective measures and subjective pilot ratings. Pilots were asked to rate agent behavior for the described use cases after each mission. The questionnaires were realized using Likert scales ranging from 1/negative to 7/positive, with 4 being neutral. Results will be discussed in the following subsections.

6.1 Use Case: Initial Planning

By default, during initial mission planning, the agent was in the *silent mode*, generating no proposals, since time and workload are no crucial factors in this situation. In case of a detected workload peak, the agent was set into the *incremental planning mode* by the workload adaptive associate system component. In 70% of the initial planning cases, the agent remained in the *silent mode* and did not take any initiative (Fig. 5).

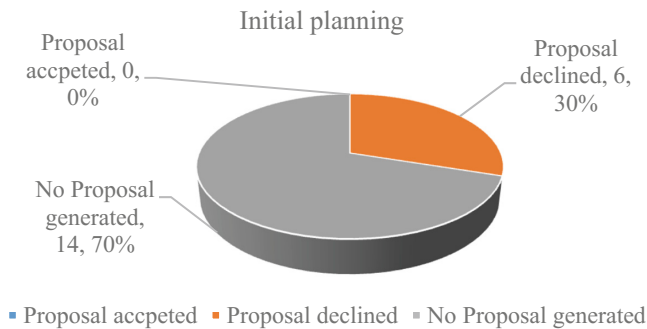


Fig. 5. Results of accepted and declined proposals during initial mission planning phase (N = 20)

In all other cases, the agent’s automation level was changed to *incremental planning mode* due to higher workload. In these cases, the agent initiated at least one incremental proposal. These proposals were either ignored or actively rejected by the subjects.

However, the extent of interventions for initial planning was rated well in general ($M = 0.2$, $SD = 0.75$), (Fig. 7). Subjects stated that they did not like incremental proposals during the initial planning process, once they started the planning process. One pilot stated that he preferred the *extensive planning mode* (ideally a proposal for a full mission plan) at the beginning of this planning phase. The other pilots preferred the *silent mode*. These results indicate, that the silent mode is the best suitable default mode for initial planning. In case of high workload, the agent could directly switch into the *extensive planning mode*. However, more research is required here.

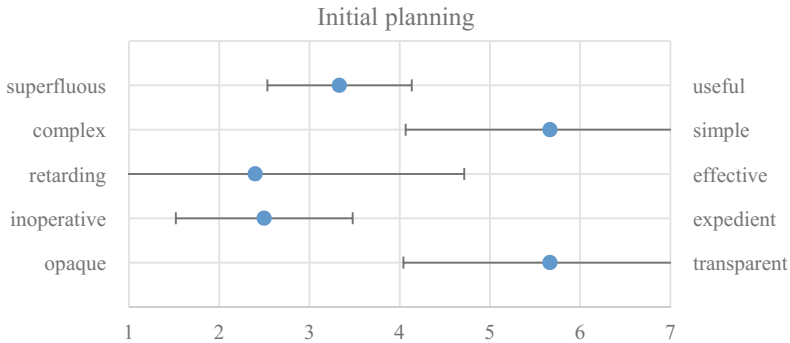


Fig. 6. Results of subjective ratings for interventions during initial mission planning (mean and standard deviation, N = 6)

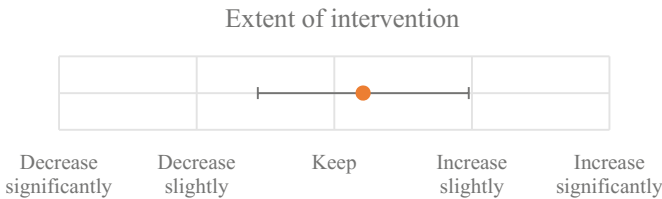


Fig. 7. Results of subjective ratings for extent of initiative for initial mission planning (mean and standard deviation, N = 20)

Results for subjective ratings for the agent intervention behavior during the initial planning can be seen in Fig. 6. The figure shows mean value and standard deviation. The figure shows subject’s ratings only for the 6 cases, when the agent generated a proposal. In general, these interventions during initial planning were rated rather unnecessary and retarding. This result confirms objective measured results.

6.2 Use Case: Re-planning Due to Threat

The second described use case comprises the change of the tactical situation due to a pop-up threat on the primary helicopter route in flight. The agent’s default mode in flight was the incremental planning mode. In such cases the agent immediately calculated an alternative route, which minimized the deviation from primary flight plan and proposed the solution to the pilot. Additionally, pilots could switch to the pre-planned alternative route (Fig. 2) or re-plan the route manually. Usually, subsequent to the re-planned helicopter route, a UAV had to be re-assigned to the new helicopter route for reconnaissance reasons. If the pilot did modify the UAV task within reasonable time (approximately 30 s after the first one), the agent generated a second proposal for this issue. Results show that 13 out of 18 proposals were accepted. Three times the pilot preferred a complete rerouting using the preplanned alternative route. In two cases, the pilot preferred a manual rerouting (Fig. 8). After half of the conducted missions, pilots were asked to rate the re-planning proposals generated by the planning agent. It can be seen that the interventions were accepted well in general (Fig. 9). The figure shows mean value and standard deviation. In a number of cases, subjects stated that the extent of intervention should be increased ($M = 0.40$, $SD = 0.70$) (Fig. 10).

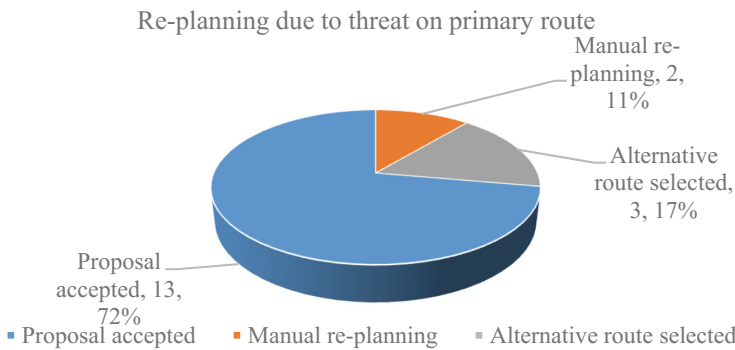


Fig. 8. Results of accepted and declined proposals during re-planning due to threat (N = 18)

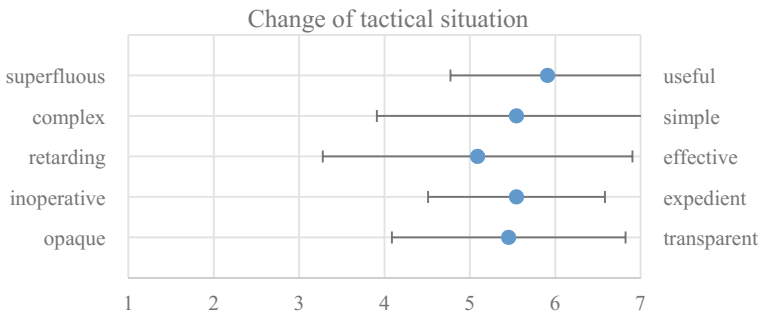


Fig. 9. Subjective ratings for agent behavior on threat detection on primary helicopter route (mean and standard deviation, N = 11)

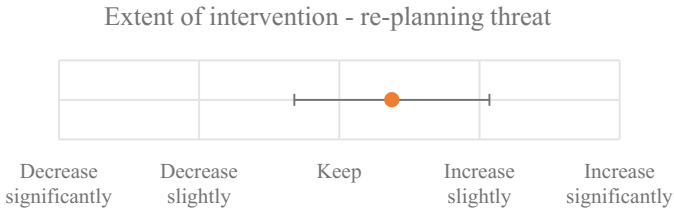


Fig. 10. Results of subjective ratings for extent of initiative for re-planning due to threat (mean and standard deviation, N = 11)

In these cases, the pilots mentioned that they would prefer the option to correct both, helicopter route and UAV task, using a single dialog.

6.3 Use Case: Re-planning Due to Change of Mission Objective

When the mission objective changed throughout a mission, the automation level of the planning agent was changed to the *extensive planning mode* due to an assumed workload peak. In this case, the agent generated a complex proposal comprising a helicopter route and a corresponding UAV task assignment for the two most important tasks (landing point detection and reconnaissance of primary route). On acceptance, the planning agent implemented these tasks automatically. Figure 11 shows the acceptance rates of proposals. In 19 use cases, (95%) the planning agent generated an extensive proposal. In 90% of all cases the proposal was accepted. The results show the high value of this type of intervention in the given situation.

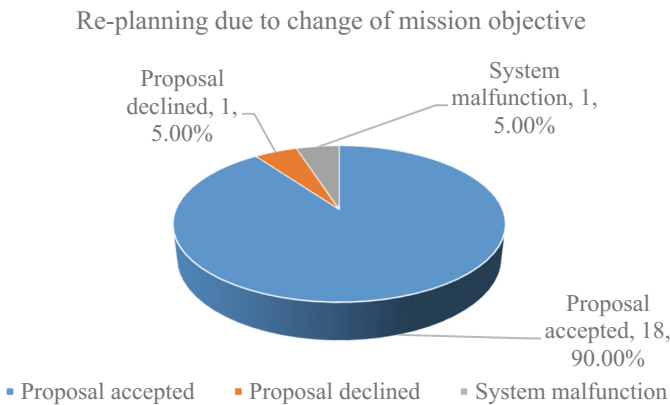


Fig. 11. Results of accepted and declined proposals immediately after change of mission objective (N = 20)

Figure 12 shows the subjective ratings for the agent behavior within the given re-planning use case. The overall ratings were very positive. The interventions were

rated helpful and effective. However, subjects also stated that the transparency of interventions could be increased. The extent of the re-planning proposals was rated very well ($M = 0.15$, $SD = 0.49$) (Fig. 13).

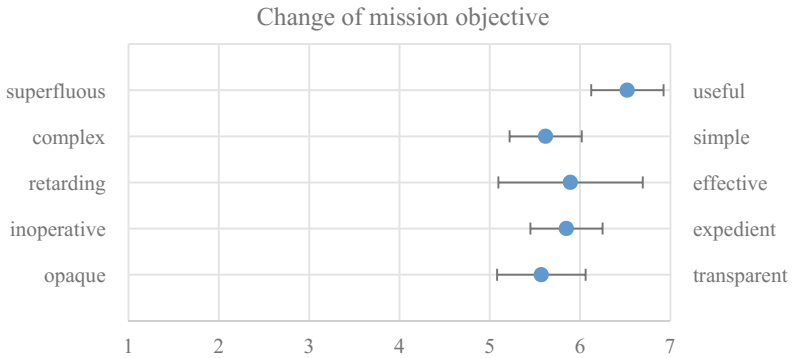


Fig. 12. Ratings for assistance after change of mission objective (mean and standard deviation, $N = 20$)

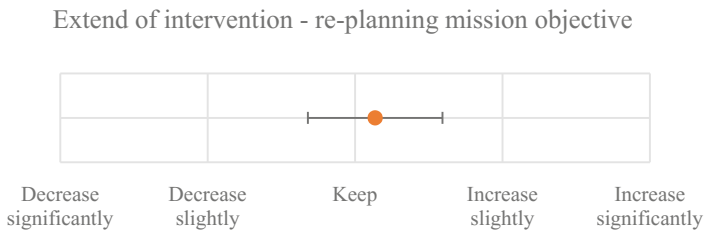


Fig. 13. Results of subjective ratings for extent of initiative for re-planning due to change of mission objective (mean and standard deviation, $N = 20$)

6.4 Overall Rating

After completion of the last mission, pilots had to rate the overall *mixed-initiate* agent behavior (Fig. 14). The figure shows that these ratings were very positive. Pilots stated that the behavior is expedient and useful. However, pilots also stated that transparency of the agent’s proposals could be increased. Several planning proposals were rather difficult to understand. This was especially the case, when the agent recommended an action, which was not visible to the pilot in the current map section. In these cases, pilots felt that they were interrupted in their work flow. Overall ratings comprise all interventions generated by the agent throughout the experimental campaign. Besides previously discussed types of intervention, this rating also includes agent task proposals for new UAV tasks.

	1	2	3	4	5	6	7	
superfluous					•	•	••	useful
inoperative					•	••	•	expedient
complex			•		•••			simple
ineffective				•	•	••		effective
opaque				•	••	•		transparent
slow				••	••			fast
insistent			•	•	••			restraint
as expected			••	•	•			surprising

Fig. 14. Overall ratings

7 Conclusion

In this article, we presented a concept for a *mixed-initiative* planning associate, which is designed to support the crew of a manned helicopter managing a MUM-T mission with up to three UAVs. Our scalable approach offered assistance on different levels of automation, depending on the tactical situation and the workload of the operator. Thereby, workload was determined by an external workload component.

We designed a full-mission experiment with crews of German army aviators. So far, we focused on the pilot’s subjective feedbacks regarding the behavior of the planning agent. Results indicate that the *scalable mixed-initiative* approach was accepted in general very well. When looking into the details, ratings could be categorized based on the three use cases. Agent interactions, generated during the initial mission planning on ground, were rated rather superfluous. Pilots would like to have either no proposals or an extensive planning proposal, which can be modified afterwards through the pilot. Agent interventions for re-planning support on changes of the tactical situation were received very well. A particularly high rating was noticeable for agent behavior after changes of the mission objective. The varying extent of intervention was received very well for all three use cases. This shows the huge advantage of our *scalability* concept in the *mixed-initiative* approach.

The evaluation of objective data obtained through the described experiment is ongoing. Future work will comprise *mixed-initiative* planning with multiple human users.

References

1. Strenzke, R., Schulte, A.: Design and evaluation of a system for mixed-initiative operation. *Acta Futura* 5, 83–97 (2012)
2. Clare, A.S., Macbeth, J.C., Cummings, M.L., Member, S.: Mixed-initiative strategies for real-time scheduling of multiple unmanned vehicles. In: American Control Conference, pp. 676–682 (2012)

3. Bresina, J.L., Jonsson, A.K., Morris, P.H., Rajan, K.: Mixed-initiative planning in MAPGEN: capabilities and shortcomings. In: Proceedings of the ICAPS–2005 Workshop on Mixed-Initiative Planning and Scheduling, p. 8 (2005)
4. Allen, J., Ferguson, G.: Human-machine collaborative planning. In: NASA Planning and Scheduling Work (2002)
5. Schulte, A., Donath, D.: A design and description method for human-autonomy teaming systems. In: Advances in Intelligent Systems and Computing, pp. 3–9 (2018)
6. Schmitt, F., Roth, G., Schulte, A.: Design and evaluation of a mixed-initiative planner for multi-vehicle missions. In: Harris, D. (ed.) EPCE 2017. LNCS (LNAI), vol. 10276, pp. 375–392. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58475-1_28
7. Simon, H.A.: Rational choice and the structure of the environment. *Psychol. Rev.* **63**, 129–138 (1956)
8. McDermott, D., Ghallab, M., Howe, A., Knoblock, C.: PDDL-the planning domain definition language. In: AIPS 1998 Plan (1998)
9. IBM Corp., IBM: V12. 1: User’s Manual for CPLEX. *Int. Bus. Mach. Corp.* **12**, 481 (2009)
10. Honecker, F., Brand, Y., Schulte, A.: A task-centered approach for workload-adaptive pilot associate systems. In: Proceedings of the 32rd Conference of the European Association for Aviation Psychology – Thinking High AND Low: Cognition and Decision Making in Aviation, Cascais, Portugal (2016)
11. Brand, Y., Schulte, A.: Model-based prediction of workload for adaptive associate systems. In: Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1722–1727 (2017)
12. Schmitt, F., Roth, G., Barber, D., Chen, J., Schulte, A.: Experimental validation of pilot situation awareness enhancement through transparency design of a scalable mixed-initiative mission planner. In: Karwowski, W., Ahram, T. (eds.) IHSI 2018. AISC, vol. 722, pp. 209–215. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73888-8_33