2015

Experimental Evaluation of Varying Feedback of a Cognitive Agent System for UAV Mission Management

Elisabeth Denk

Sebastian Clauss

Annike Borchers

Josef Werner

Axel Schulte

EXPERIMENTAL EVALUATION OF VARYING FEEDBACK OF A COGNITIVE AGENT SYSTEM FOR UAV MISSION MANAGEMENT

Elisabeth Denk, Sebastian Clauss, Annike Borchers, Josef Werner & Axel Schulte
Universität der Bundeswehr München (UBM)
Institute of Flight Systems (LRT- 13), 85577 Neubiberg, Germany
{elisabeth.denk, sebastian.clauss, axel.schulte}@unibw.de

In this study we investigate on a cognitive delegation agent for UAV task-based mission management. Particularly, we advocate a specific high-level feedback provided by the agent to the human operator to enhance mission effectiveness. As extension to human supervisory control we suggest to introduce the concept of agent supervisory control where the agent is delegated by high-level operator commands and controls several sub-systems aboard the UAV fulfilling the mission in highly automated fashion. Results of our experimental human-in-the-loop study focusing on the effects of high-level feedback are presented. Therefore, two configurations are compared, one with basic feedback and one with full feedback. The results show that particularly during in-flight re-planning situations the full feedback is beneficial. Interaction times and task related activities are significantly lower. In the pre-flight mission preparation phase no significant effect were found. Results can be used to further develop highly automated (multi)-UAV mission management systems.

Agent Supervisory Control

Nowadays, military *Remotely Piloted Aircraft System* mission management is in focus of research (Clauss & Schulte, 2014; Theissing & Schulte, 2015). In modern UAV-systems, conventional automation (i.e., auto-flight systems) relieves the operator of high-frequent sensor-motor tasks and improves precision and performance for mission execution. Instead of manual control, the operator controls the aircraft intermittently through automation. The automation has to be monitored more or less continually by a *Human Supervisor* (HS). This type of control relationship was described as *Human Supervisory Control* (HSC) by Sheridan (1992).

The HS generally performs five *Supervisory Functions*. Determining the current objective and exploring a strategy to achieve it, using the given means (*plan).* The HS conveys its commands to the automation (*teach*) and monitors the automation to ensure proper execution (*monitor*). If necessary the HS intervenes (*intervene*) and finally may learn from experience to perform better next time (*learn*) (Sheridan, 1992). The cognitive capabilities of the HS, allow the overall system to react to individual challenges in the environment and the status of the UAV-system and enable it to compensate for unforeseen events.

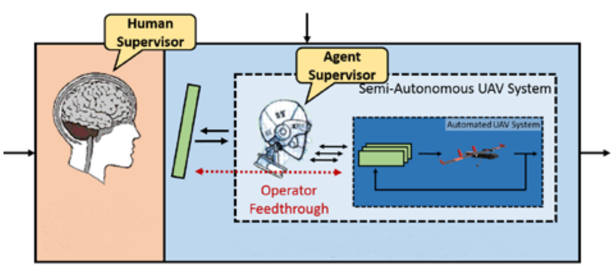

Figure 1. Work system of a semi-autonomous UAV with cognitive agent.

In this context, feedback information must be perceived, interpreted and processed by cognitive functions. To extend the operators support, a rather conventional approach of introducing more complex automation could constrain his work; increase the complexity of the overall system as well as the number of automation functions.

In this context, Bainbridge (1983) describes two *Ironies of Automation*, the first is a shift of human errors from manual control to designing and implementing of automation functions and the second is *Clumsy Automation* (Wiener, 1988). It supports the operator in low-stress situations, but cannot provide support in highly intense situations. For the supervision of automation functions in manned flight, Billings (1997) describes four *Costs of Automation*: *complexity*, *brittleness*, *opacity* and *literalism*.

To tackle some of these issues, cognitive capabilities are displaced such as decision-making, problem solving and planning aboard the aircraft. A cognitive agent, implementing cognitive capabilities, is introduced onboard the UAV to manage and control its existing conventional automation systems. In terms of *Cognitive Automation* (Onken & Schulte, 2010), the agent works within the supervision of the human operator and thus serves as a link between the mission management layer in the responsibility of the human pilot and the mostly automated UAV navigation, guidance and control.

Figure 1 shows the resulting work system from integrating a cognitive agent into the automated UAV-system. The human operator interacts with the single cognitive agent, rather than the multitude of automation functions. The agent supervisor formulates discrete commands for the conventional automation and monitoring their execution. The cognitive capabilities of the agent allow deriving action plans from human delegated objectives. For this purpose, the agent plans and coordinates the application of the underlying automation. Still, the semi-autonomous agent does not have authority to modify or specify its own objectives (Onken & Schulte, 2010). In analogy to the definition of HSC, the relationship between the cognitive agent, the conventional automation and the UAV-system may be best described by the term *Agent Supervisory Control* (Clauss, Kriegel, & Schulte, 2013). But the cognitive agent is always acting like an *intelligent* subordinate to the human within the concept of HSC, using its cognitive capabilities to execute human tasks in a flexible manner.

Figure 2 shows the resulting management hierarchy of the UAV-system including, the additional echelon as the guidance layer of the cognitive agent. The agent is introduced between the operator and the conventional automation. In this role the cognitive agent combines commanding and monitoring as well subordinating to the human operator. The agent supervisor behavior and the interaction with the human will have a combined effect on the operator's perception.

Agent Feedback within a Task-Based Guidance Approach

The human pilot acting as a supervisor of the semi-autonomous UAV-system requires information, which allows monitoring its performance and to intervene (re-plan) when necessary. Additionally, the human behavior and criteria for delegating the cognitive agent are depending on the operator's information about the agent's capabilities and performance. So the operator has to decide, which tasks must be delegated and which could be done manually (Leana, 1986; Parasuraman & Riley, 1997). For supervisory control of conventional automated systems, sophisticated concepts of interactions already exist. In the following we examine an approach to a bidirectional information flow. This allows a calibrated delegation of tasks to the cognitive agent as well as it provides adequate feedback to the operator.

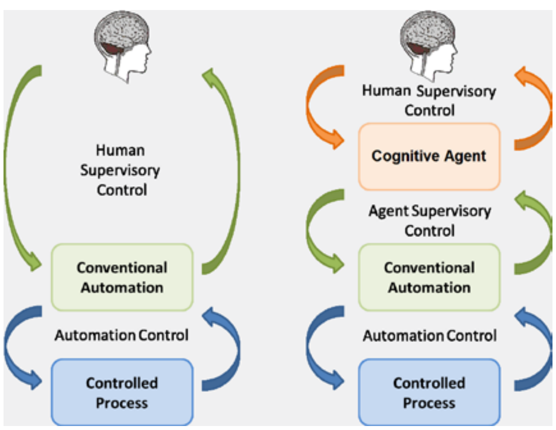

Figure 2. Agent Supervisory Control (ASC) as additional guidance loop between operator and conventional automation.

As a concept of delegation we chose a Task-Based UAV Guidance (TBG) approach (Uhrmann & Schulte, 2011). Herein the operator solely defines the objectives for the agent as commanded intents, instead of formulating step-by-step instructions for multiple automation components. Therefore, the operator defines what the semi-autonomous system shall accomplish, instead of providing *how* (i.e. through what actions) this shall be achieved. Tasks might be military reconnaissance missions. TBG stems from an inter-human delegation relationship and relieves the supervisor from the tedious task to derive automation action instructions from intentions (Clauss & Schulte, 2014).

With respect to the operator's tasks, the performance of the system is mainly affected by its ability to decide which tasks to delegate to the agent and how to formulate these tasks. Parasuraman and Riley (1997) presented criteria for the task delegation to subordinate automation (cognitive agent) and indicates those criteria that can be directly influenced by automation design. The central criterion is *reliance*, resembling the affinity of the operator to delegate a task. The human reliance on automation is directly influenced by the *confidence* in a manual executing of the task, the current level of *fatigue*, the *perceived risk* associated with task failure and *trust in automation* for a satisfying task executing.

Three specific criteria in this context (*machine accuracy, trust in automation and workload*) can be identified as affected directly by automation behavior and accordingly at least to some extent controllable by design. Machine accuracy describes the level of sufficiency with which a delegated task is executed by the automation. Trust in automation comes from the operator's perception of the of automation behavior and is used to predict behavior in future situations. The operator's sensor-motor and cognitive workload is directly influenced by the interaction with automation during task execution. The likelihood of trust in task delegation depends on the complexity of the underlying system (Lee & See, 2004). The basis for trust development, regarding the agent's capabilities, is the information feedback received by the operator. This information can be categorized (Lee & Moray, 1992). The operator's monitoring task is accomplished with the help of the assessment of agent feedback. The agent itself monitors the heterogeneous conventional automation systems and creates a symbolic representation, which it communicates to the human. With regard to the operator's workload, the human is supported by the agent, if the cognitive task (interpreting the symbolic information) is less complex than the processes needed to monitor the conventional automation itself. The desired form of feedback, with respect to its application domain, can be described by the term *etiquette* (Miller, 2002). Etiquette means an established form of interaction that expresses the role of the transmitter and which rules aim a better understanding and enhancing the effectiveness and the safety of the communicating system.

An extended feedback of a cognitive agent was developed (see figure 3 right side, *Enhanced Feedback Configuration*) to minimize mission errors and to raise human awareness about UAV resources and capabilities. The cognitive agent provides event-independent information about the current system status of the UAV, the current tactical situation information and perception results (threats, tactical elements) and the status and the current objective of the agent, its (reviewed) task-agenda and its execution progress. Further, the agent provides information only on currently available automated capabilities. As a feedback to the task-agenda delegated by the operator, the agent presents its execution plan containing a list of activities to be performed in order to transition from its current state into the specified goal state (Agent Plan). In case of plan execution errors, the agent uses its knowledge-based reasoning to independently derive an alternate action plan. If no solution can be derived within the boundary conditions specified by the operator, the agent reports an error reason to the operator. Figure 3 shows the map displaying a mission for each configuration. Figure 3, right (i.e., *Enhanced Feedback Configuration*) shows the task agenda (reviewed by the cognitive agent). The blue line indicates the flight plan created by the agent.

Figure 3, left (i.e. *Baseline Configuration*) shows the setup where the agent's feedback is limited to only graphical information about the position of the UAV and its current flight plan. In this configuration the agent is fully functional, but no information about its intent or task execution plan is conveyed to the operator. Tasks may be delegated by the operator with no visibility of the UAV's actually available capabilities.

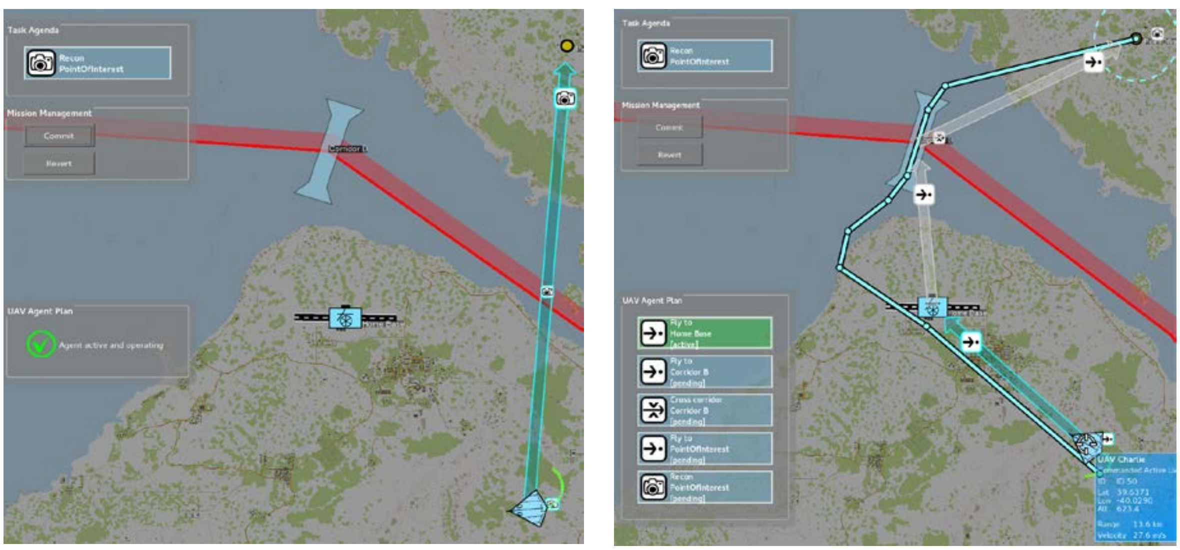

Figure 3. Moving map and planning display in mission and payload control station (left: *Baseline Configuration*, right: *Enhanced Feedback Configuration*).

Experimental evaluation

Communication between the operator and the automation on a symbolic level is an essential part of task based guidance. We hypothesize that the feedback provided by the agent will affect the work result of the system, even if its decision-making and control functions stay unmodified. In an experimental campaign the impact of the agent's feedback behavior on planning, commanding and re-planning of the operator is examined. The developed cognitive agent (Clauss & Schulte, 2014) with advanced feedback abilities is evaluated with respect to human performance, re-plan capability and human trust in the agent. In our experiment we compare the full feedback system to a configuration with reduced feedback information.

Research setup and configurations

The experimental design is a within-subject design with a secondary task (Borchers, 2014; Werner, 2014). Its factor is the feedback behavior of the cognitive agent during mission execution (*configuration A* and *B*). For its evaluation, a comparative experiment was conducted, in which two missions (m*ission I* and *II*) are performed. Both missions are very similar, so they are comparable (see experimental procedure). The participants perform *mission I* and *II* while agent configuration A (*Baseline Configuration*) and separately with agent configuration B (*Enhanced Feedback Configuration*) (see figure 3). All participants were exposed to the two configurations of the system, while performing missions and completing the questionnaires. To eliminate sequence effects and spillover effects, the missions were randomized.

The experimental hypothesis says that configuration B reduces workload, planning effort and error rate, compared to configuration A. It can also be assumed, that configuration B leads to higher situation awareness, distribution of attention, trust in automation, visual perception and acceptance, as compared to configuration A. In order to prove the hypotheses the constructs were operationalized using the following dependent measures.

Objective performance measurements examine the objective performed tries for planning or re-planning a mission. An additional performance variable is the amount of the error rate while planning and re-planning. Human-system interactions were measured to determine behavior changes. Therefore, the interaction time and the interaction activity were counted. The interaction time is defined as time where the operator is actively interacting with the cognitive agent in a particular situation. The interaction time can be quantified by the observation of manual actions (clicks) on the touch displays or by use of eye tracking data during monitoring tasks. The interaction activity is measured with the number of touch-screen clicks over time. Subjective dimensions were used to measure the subjective workload and performance by the standardized NASA-TLX (Hart, 1986; Hart & Staveland, 1988). The six-scale questionnaire was answered after planning and commanding, monitoring and re-planning in both missions. Additionally to the performance measuring, to observe the situation awareness of the operator the SAGAT questionnaire (Endsley, 1988) was performed after re-planning. At the end of both missions the operator had to complete a questionnaire for the subjective evaluation of acceptance and trust in automation (Lee & Moray, 1992; Lee & See, 2004). The questionnaire consists of four dimensions (system interaction, system behavior, system information and overall system).

Participants and experimental conditions

The sample consists of 13 officers of the German Armed Forces. The participants aged between 21 to 27 years (M_{age}=24.2) are recruited from the University of the Bundeswehr Munich. Participants include 12 male and one female student.

The operators control the UAV-system from a *Ground Control Station* (GCS), using a Human-Machine Interface (HMI) consisting of two multi-touch displays. Inputs can be made using touch screen or mouse controls. Room dividers shield the GCS to avoid visual distractions for the operators, while a headset damps external sounds and facilitates intercom. The lower screen features the MPCS (Theissing & Schulte, 2015) for task-based UAV guidance and automation monitoring using a map display of the mission area and a graphical representation of the tactical situation. The interaction with mission elements allows the formulation of tasks provided to the UAVs. The upper screen of the GCS features a modular sensor interface showing the live sensor-feed for the UAV. The sensor display is used to observe the surveillance area and to detect vehicle movements. The GCS experimental setup includes an eye-tracking system, measuring the operators' focal point on the screens during the experiments. The experimenter uses an external workstation for the manipulation and control of the experiment's tactical situation and events (Borchers, 2014; Werner, 2014).

Experimental procedure

During the experiment, each participant performed *mission I* and *mission II*. The experiment has a total duration of approximately three hours, including the mission preparation, the actual missions and standardized briefings and debriefings.

The missions have an identical general layout with similarly complex mission scenarios but differing mission events and dynamic threats. The sample is split into two, whereas the first half performs *mission I* in system configuration A followed by *mission II* in system configuration B. The second half performs both missions with switched system configurations. The assignment to the conditions was randomized to minimize systematic effects.

Each subject performs two consecutive missions (*mission I* and *mission II*), each with similar mission layout and tasks. An island, conquered by hostile forces, is to be retaken from an adjacent island and therefore periodic reconnaissance missions have to be performed by the operator and his UAV. In this scope, own troops should be supported, hostile targets detected and identified as well as areas scanned (reconnoitered). At the beginning of each mission the UAV takes off from its home base (indicated by the blue square below the UAV symbol) and crosses the FLOT (red line) through transit corridors (blue). Enemy *Surface-to-Air Missile Sites* (*SAM-Sites*), indicated in red, generally have to be avoided by the UAV. Reconnaissance targets are objects or areas (yellow). The main objective for *mission I* is to perform a detection task in two areas (A, B) and a reconnaissance in two additional areas (C, D). The operator should plan and complete the mission. After completing the main objective the mission is interrupted and the operator has to re-plan. In *mission II*, area A and B should be detected and area C cleared. During the mission positions are also changing and the UAV must land on an alternative airfield (re-planning). The procedure and all instructions are standardized.

Experimental Results

The explorative hypotheses are calculated with the Wilcoxon signed-rank test. For the exploratory data analyses, the significance level was set to 5%.

While planning and commanding the mission, there are no differences between configuration A and B in the following variables. The interaction time is in both configurations not significant different (Z= -0.629; p=.277), as well as the interaction activity (clicks) (Z= -0.594; p=.294). The data of the subjective workload shows no significant differences between configuration A and B (Z= -0.874; p=.207), as well as configuration A to baseline data (Z= -0.314; p=.395) and configuration B to baseline data (Z= -0.735; p=.242). Figure 5 presents the results of the number of clicks and interaction time, while planning and commanding.

While monitoring the mission, in configuration A (MW = 4.24, SD = 3.79) the interaction time is significant shorter than in configuration B (MW = 6.04; SD = 3.32; Z= -1.784; p=.042). The longer interaction time during configuration B during the monitoring phase might be an indication for a higher task load. But the higher task load does not cause a higher subjective workload (Z= -0.559; p=.300). The operator must perceive more information, which needs more time. An explanation for no differences between configuration A and B might be that a different interaction behavior does not exist because of the short latency, till planning and commanding the mission.

While re-planning the mission, in configuration A (MW = 2.69; SD = 1.03) the interaction activity of re-planning is significant higher than in configuration B (MW = 1.67; SD = 0.98; Z= -1.895; p=.045). The error rate however is significant higher in configuration A (MW = 2.15; SD = 0.90), as compared to configuration B (MW = 0.62; SD = 0.96; Z= -2.831; p=.001). Additionally, in configuration A (MW = 68.46; SD = 51.31) significant more clicks are made than in configuration B (MW = 23.00; SD = 21.61; Z= -3.059; p=.000). The interaction time in configuration A (MW = 112.60; SD = 58.38) is significantly higher than in configuration B (MW = 50.03; SD = 44.54; Z= -2.432; p=.006). Compared to the baseline data (MW = 17.15; SD = 8.46), in configuration A (MW = 38.79; SD = 17.13) the subjective workload is significant higher (Z= -3.059; p=.000). In configuration B (MW = 37.86; SD = 13.00) the workload is also higher than in the baseline (Z= -3.182; p=.000), but between configuration A and B are no significant workload differences (Z= -0.105; p=.473). The situation awareness after re-planning in configuration A and B (Z= -0.735; p=.236) is equal. Figure 5 presents the results of the number of clicks and the interaction time for the re-planning phase.

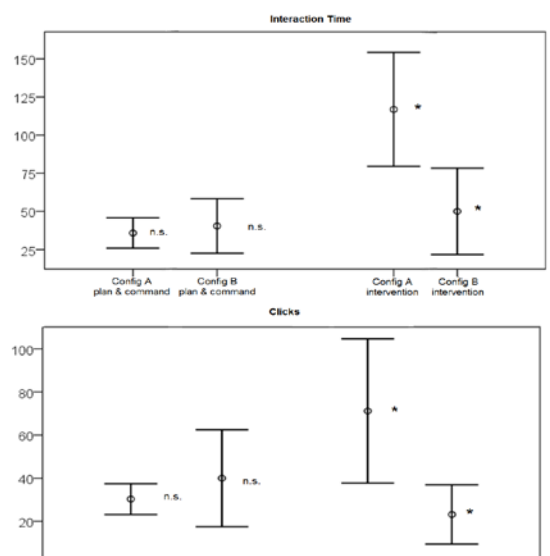

Figure 5. Result of the interaction time and the number of interactions during planning and commanding as compared to re-planning in configurations A and B (n.s. = Not significant, * = $p<0.01$).

The evaluation shows that in general there are differences in the subjective ratings between the two configurations. Furthermore, configuration A is less accepted than configuration B in the subscales system behavior (configuration A: MW = 3.95; SD = 1.46; configuration B MW = 4.70; SD = 1.12; Z=-2.473; p=.005) and overall system (configuration A: MW = 112.60; SD = 58.38; configuration B MW = 112.60; SD = 58.38; Z=-2.665; p=.002). The operators' trust in automation in configuration A (MW = 4.44; SD = 1.08) is lower than in configuration B (MW = 5.04; SD = 0.77; Z=-2.518; p=.004), because the system acts like the operator requests. This might be a reason why the overall system itself is more accepted.

Conclusions

This paper describes an approach to cognitive agent feedback and its experimental evaluation concerning human delegation behavior. The developed cognitive agent is designed to support the human operator to execute supervisory control functions in a highly automated technical system.

While no effects of the provided feedback could be identified for the initial planning phase of the mission, the re-planning phase is positively affected by the provided agent feedback. More feedback information requires more processing time of the human operator while monitoring the system, but raises overall mission efficiency (lower error rate) and lowers the operator's interaction and failure rate during re-planning.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19* (6), 775–779.

Billings, C. E. (1997). *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: LEA

Borchers, A. (2014). *Konzeptionierung und Durchführung einer Versuchsreihe zur Evaluierung eines kognitiven Agenten in Bezug auf die Planung und Kommandierung von Aufklärungsmissionen. [Configuration and performance of a research series, to evaluate a cognitive agent, while planning and intervention of a reconnaissance mission] (*Master Thesis). Universität der Bundeswehr, München, Germany.

Clauss, S., Kriegel, M. & Schulte, A. (2013). *UAV Capability Management using Agent Supervisory Control*. In: AIAA Infotech@Aerospace Conference 2013. Boston, Massachusetts, USA. 19-22 August 2013.

Clauss, S. & Schulte, A. (2014). *Implications for operator interactions in an agent supervisory control relationship*. In: International Conference on Unmanned Aircraft Systems (ICUAS). Orlando, FL. 27-30 May 2014

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Human Factors and Ergonomics Society Annual Meeting, 32* (2), 97–101.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload*. NL: Elsevier.

NASA (1986). Task Load Index (TLX): Computerized version (Version 1.0). Moffett Field, CA: Human Research Performance Group, NASA Ames Research Center.

Leana, C. R. (1986). Predictors and Consequences of Delegation. *TAMJ, 29* (4), 754–774.

Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *JoHFES, 46* (1), 50–80.

Lee, J. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics, 10* (35), 1243–1270.

Miller, C. A. (2002). *Definitions and dimensions of etiquette*. North Falmouth, MA: AAAI Fall Symposium on Etiquette for Human-Computer Work.

Onken, R. & Schulte, A. (2010). *System-ergonomic design of cognitive automation: Dual-mode cognitive design of vehicle guidance and control work systems*. Berlin, Heidelberg: Springer-Verlag.

Parasuraman, R. & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *HF, 2* (39), 230–253.

Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA: MIT Press.

Theissing, N., & Schulte, A. (2014). *Flight Management Assistance through Cognitive Automation Adapting to the Operator's State of Mind*. In: 31st EAAP Conference. Valetta, Malta. 22-26 September 2014.

Uhrmann, J. & Schulte, A. (2011). *Task-based Guidance of Multiple UAV Using Cognitive Automation*. In: The Third International Conference on Advanced Cognitive Technologies and Applications. COGNITIVE 2011. Rome, Italy. 25-30 September 2011.

Werner, J. (2014). *Konzeptionierung und Durchführung einer Versuchsreihe zur Evaluierung eines kognitiven Agenten in Bezug auf die Überwachung von Aufklärungsmissionen. [Configuration and performance of a research series, to evaluate a cognitive agent, while monitoring a reconnaissance mission] (*Master Thesis). Universität der Bundeswehr, München, Germany.

Wiener, E. (1988). Cockpit Automation. In E. Wiener (Ed.), *Human Factors in Aviation* (pp. 433–462). CA: AP.