Retrospective Verbal Probing in Evaluation of Pilots' Situation Models in Simulated Air Combat

Heikki. P. Mansikka[1,2], Kai M. Virtanen[1,3], and Don Harris[4]

[1]Department of Mathematics and Systems Analysis, Aalto University, Helsinki, Finland

[2]Insta DefSec, Tampere, Finland

[3]Department of Military Technology, Finnish National Defence University, Finland

[4]Faculty of Engineering, Environment and Computing, Coventry University, Coventry, UK

Abstract. For a fighter pilot to be effective, his/her situation models (SMs) should reflect the objective reality of air combat. The difference between the pilot's knowledge about the SM related concepts and an approximation of the objective reality was used as an indirect measure of pilot's SMs' accuracy and correctness. The utility of the retrospective verbal probing (RVP) technique was analyzed in the evaluation of SMs. Twenty-eight F/A-18 pilots flew a simulated air combat mission where the task complexity was manipulated. The results obtained with the RVP technique consistently reflected the manipulation of the task complexity and it avoided most of the typical challenges related to the use of freeze-probe, observer and self-rating techniques in the evaluation of SMs.

Keywords: air combat, situation models, measurement, virtual simulation, retrospective verbal probing

Introduction

Air combat is a complex and dynamic system. Understanding and controlling such a system can be cognitively demanding. The pilots' task in an air combat is to keep themselves in an offensive position that increases the probability of weapon intercept with the enemy, while simultaneously denying or lowering the enemy's probability of achieving the same. To achieve this, pilots need to understand the components of the system, to identify the interconnections, delays and feedback mechanisms between the system components, and to predict how the system behaves over time (Sweeney & Sterman, 2000). In other words, pilots' situation models (SMs) (Endsley, 2000) should be closely aligned with the architecture and components of the system, as accurate and correct SMs are essential contributors to high performance and operational effectiveness (Endsley, 1995a). Therefore, a great deal of fighter pilot training is about evaluating how accurately and correctly their SMs reflect the objective reality of air combat.

Pilots' SMs are dynamically updated mental models (MMs) which represent their knowledge and understanding of the current and near-future states of the system they are interacting with (Wickens et al., 2004). MMs, on the other hand, are a collective name for the structure and content of a person's understanding regarding the elements within their environment and the sequence of activities regarding the task (Gilbert, 2011; Wilson & Rutherford, 1989). MMs are comprised of schemas and scripts held in the pilot's long-term memory (Moray, 1998). Schemas describe the structure and content of the pilot's knowledge, whereas scripts depict the sequence of activities and behaviors related to a specific task (Wilson and Rutherford, 1989). Both scripts and schemas are context specific in the sense that each is associated with a particular concept (Langan-Fox, Code, & Langfield-Smith, 2000) within the pilot's environment (Wickens et al., 2004). Endsley (1995a) considers SMs as a hierarchical state of situation awareness (SA) with three distinctive levels; perception (SA level 1), comprehension (SA level 2) and prediction (SA level 3). Although Endsley uses SA and SMs as synonyms (Endsley, 2000), this study uses terms SMs and SM levels instead of SA and SA levels – mainly to emphasize the role of MMs and the way they are dynamically updated with observations.

During an air combat mission, pilots have SMs concerning concepts such as friendly and enemy aircraft's locations, parameters and capabilities; environment, geography and airspace; ground and surface forces; enemy behaviors and maneuvers; and friendly team's tasks and objectives. The pilots often utilize several SMs concurrently. When the pilots' SMs and the objective reality are closely aligned, the pilots can understand how the different air combat concepts interact and how the system as a whole operates (Stout et al., 1999). The evaluation of SMs seeks to determine how accurately and correctly the pilots' SMs reflect the objective reality of air combat. While SMs cannot be measured directly (Kraiger & Wenzel, 1997), they can be evaluated indirectly by measuring the pilots' knowledge about the concepts associated with SMs and by determining how closely the knowledge is aligned with the ground truth. The ground truth is an approximation of the objective reality and represents the real state of the SM related concepts as opposed to the pilot's knowledge affected by interpretation and inference.

The term "elicitation" is often used in the context of SMs. Elicitation basically refers to the process of inquiry to encourage a person to externalize their knowledge about the concepts (Jones et al., 2014). Put simply, elicitation is essentially about giving the person's knowledge a form, and then measuring it. Techniques to measure the pilots' knowledge about the concepts can be broadly categorized as either performance techniques, real-time probe techniques, freeze-probe recall techniques, self-rating techniques or observer rating techniques (Salmon et al., 2009). Each of the measuring techniques has limitations, especially when applied to air combat. For example, when performance measures are used, pilots tend to attribute their low performance with subjectively assumed inadequate knowledge (Endsley, 1995b). It has also been argued that the performance measures are not necessarily measures of pilots' knowledge at all (Mansikka et al., 2019). Real-time probe techniques, on the other hand, are limited by the fact that during the intensive phases of air combat where accurate and correct SMs are most critical, pilots don't have time to react to the probes and additional tasks used in the real-time probe techniques. The utility of the freeze-probe techniques is highly limited by the fact that stopping a mission in virtual simulation is highly disruptive – and is impossible during real flight. In addition, the probes used in the freeze-probe techniques are typically highly detailed and the number of probes required to draw a comprehensive, or even representative, picture about the pilots' knowledge of a complex air combat make the data collection in an operational environment too time consuming. On the other hand, self-rating techniques are low-cost, easy to administer and non-intrusive – especially when conducted ex post facto. However, by simply asking a pilot whether his/her SMs are aligned with the objective reality is likely to result in honest, but false results. Finally, observer rating techniques ascertain the pilot's knowledge from performance; the appropriateness of the observed responses to discrete events is considered to enable the comparison of the actual responses with the expected responses (Pritchett & Hansman, 2000). Expected responses are used to form the basis for an observable measure of the pilot's knowledge regarding the concepts, as correct actions can be anticipated only through the pilot having SMs which are aligned with the objective reality. However, due to the inherent nature of observer rating techniques, they cannot reveal the covert aspects of knowledge. Despite their known limitations, however, observer rating techniques are among the few techniques suitable for air combat simulations – mainly as they are non-intrusive.

Interviews (Carley & Palmquist, 1992; Langan-Fox, 2002; Morgan et al., 2002) are alternative techniques to measure pilot's knowledge about the concepts in a fast-paced air combat simulation. If an interview is conducted shortly after the activity of interest, it is non-intrusive and can be used in a natural task setting. Moreover, where most techniques to measure pilots' knowledge about the concepts typically address narrow aspects of knowledge (e.g., altitude or speed of a target at certain moment), an interview technique allows addressing more complex knowledge structures and broader knowledge content (e.g., interaction of the system elements or operation of the system as a whole). While an interview-based approach can reveal a comprehensive picture of the pilot's knowledge about the concepts, its use is complicated by several factors. First, it is unlikely that the full spectrum of SMs and the content of a pilots' knowledge about the concepts can ever be fully captured. Second, while it is recommended that SMs are evaluated in a natural task setting, most interview techniques are time-consuming, restricting their use in air combat training (Langan-Fox et al., 2000; Langan-Fox et al., 2004).

One interview technique is a retrospective verbal probing (RVP) (Ericsson & Simon, 1980; Willis et al., 1991) technique, an application of cognitive interview (Geiselman et al., 1986). In the RVP technique, pilots are probed about the cognitive processes that occurred at an earlier point in time. Before the RVP technique can be applied, the relevant concepts and the probes addressing the pilot's knowledge about them must be identified. In that sense, the RVP technique is quite similar to most knowledge measuring techniques, where the relevant concepts are first identified, followed by the selection of events and behaviors to be observed or questioned to reveal the pilot's knowledge about these concepts.

When the RVP technique is used to measure a pilot's knowledge about the concepts in simulated air combat, the pilot flies a combat mission and later attends a debrief. As a standard procedure, a flight instructor facilitates the debrief and reconstructs the mission for a review and analysis. All data gathered from the mission are typically available in the debrief (e.g., cockpit recordings, radio communications and missile simulations). Therefore, the reconstruction is often referred to as the ground truth of the mission (Waag & Houck, 1994). The ground truth is made available to assist the pilot in recalling his/her knowledge, decisions and activities related to the mission. The application of the RVP technique starts with the pilot reviewing the mission reconstruction with the facilitator. Once the facilitator identifies an event or activity related to an individual concept, the reconstruction is paused, and the probes addressing the pilot's knowledge about the concept in question are introduced. A probe or set of probes is presented to initiate the interview. Then, the facilitator assists the pilot to externalize his/her knowledge about the concept with additional probing as necessary. The facilitator also supports the pilot's recall by pausing, playing, zooming and rewinding the mission reconstruction as needed. Once the pilot's knowledge about the concept has been externalized, the pilot's knowledge is compared with the ground truth. The difference between the pilot's knowledge and the ground truth is evaluated and scored, and it is used as an indirect measure of pilot's SM accuracy and correctness. Then, the review of the mission reconstruction is continued until an event or activity related to the next concept is identified; the mission reconstruction is paused, and the next probes are introduced to initiate the interview. The procedure is repeated until the pilot's knowledge about all concepts have been measured and compared with the ground truth.

The interview procedure used in the RVP technique is similar to a normal debrief, where the pilots routinely (but often informally) recall their knowledge about the concepts and compare it with the ground truth of the mission. Fundamentally, most air combat learning effects count on the pilots' ability to recall past events during debriefs.

In this study, pilots' SMs were evaluated in a simulated air combat mission by measuring their knowledge about selected concepts using the RVP technique, followed by a comparison of this knowledge with the ground truth. Task complexity was manipulated within the mission. It was hypothesized that the RVP technique would be sensitive to the manipulation of task complexity.

Method

Participants

Twenty-eight combat-ready F/A-18C pilots were recruited. As there were no female combat-ready pilots available, all participants were male. The participants' mean flying experience on F/A-18 was 683 hours (SD=341). All participants were fit to fly. The participants were familiar with the tactics to be employed and qualified to fly the mission profile used in the study. The trial missions were flown during normal office hours. The trial mission was part of normal flight training and at the day of the trial the participants attended other flying duties according to their training schedule. The data were collected non-invasively. Written, informed consent was obtained from each participant.

Test design

A standard fighting unit in air combat is a flight, which refers to a team of four pilots. A flight is composed of two elements, a lead element and a wing element. Both elements have two pilots each, a leader and a wingman. In this study, only the wingmen's SMs were evaluated.

Air combat operational test and evaluation pilots prepared a realistic beyond-visual-range (BVR) air combat mission. In BVR air combat, the pilots use airborne detection equipment to search for the enemy aircraft and employ remote air-to-air missiles to attack them while at the same time staying BVR from the enemy aircraft (Paddon, 1977). During preparation, a group of subject matter experts conducted pre-testing by flying the trial mission number of times and evaluated its suitability for the data collection by studying the mission playbacks.

To standardize the mission flow between the participants, all constructive (i.e., computer programmed) simulation entities were designed to follow a predefined script, ensuring that the mission evolved in a similar manner for each participant. To increase the sense of authenticity, the radio calls of all constructive simulation entities were prepared as an audio file which was then synchronized with the simulation.

The mission started from combat air patrol and continued with three seamlessly connected BVR engagements, each with a slightly different target presentation. A BVR engagement refers to an isolated attack against an air threat with a directive or authorization to use sensors and/or weapon systems against designated targets. The complexity of the engagements was designed to increase towards the third engagement by complicating the enemy presentation. The participants were not informed about this manipulation. As the engagements were designed to form a logically progressing mission, the order of the engagements was not randomized.

All engagements had the following common phases: 1) target assignment, search and identification; 2) weapon employment; and 3) evasion and egress. The next engagement commenced immediately following the completion of the evasion and egress phase of the preceding engagement.

The mission was programmed into a virtual F/A-18C flight training device (FTD). The FTD had a 135-degree visual display and a fully functional cockpit. The FTD is routinely used for basic and advanced fighter pilot training. It replicates F/A-18C flying characteristics and cockpit interface with such an accuracy that the pilots can use it to fly their annual proficiency checks.

The mission was briefed using a handout, which the participants studied for ten minutes. After that, the participants entered the FTD, prepared the cockpit, and the simulation was started. Once the simulation was initiated, it followed a pre-defined script and continued to the end of the third evasion and egress phase without stopping. The overall duration of the mission was just over seven minutes.

Measures

Based on a review of air combat manuals and research articles (Endsley, 1993; Houck et al.; RoKAF, 2005), the following concepts were identified as relevant for the designed mission: aircraft locations and flight parameters, tactics, flight's and flight members' tasks and flight's objectives. Fifteen probes (later referred to as RVP probes) were formulated to tap the pilots' knowledge about the concepts. The same probes were used in each engagement. As a result, 45 RVP probes were used in each trial. Each RVP probe where the pilot's answer was aligned with the ground truth was scored as '1', whereas each RVP probe where the pilot's answer and the ground truth were not aligned was scored as '0'. The scored responses to the RVP probes formed the RVP scores. The maximum RVP score in each trial was 45. Every RVP probe was designed to tap a concept associated with a specific SM level. Each SM level related concept was tapped with five probes, see Table 1. RVP scores grouped for each SM level are referred to as RVP level (1-3) scores. As the complex interconnections, delays and feedback mechanisms of air combat are sometimes open to interpretation, it was not always easy for the pilot to provide straightforward yes/no answers to the RVP probes. Therefore, open ended verbal questioning was used to assist the pilot to externalize the knowledge structures and content he possessed, and what the ground truth actually was like. Additional questioning was also used to ensure that the pilot actually had the knowledge and understanding he reported – instead of just claiming or wrongly perceiving so.

Table 1. RVP probes used to tap the concepts associated with different SM levels

SMs level 1	SMs level 2	SMs level 3
Did you correctly perceive your flight members' and your position with respect to the selected tactic?	Did you correctly comprehend the timeline, and your flight members' and your position within it?	Did you correctly project how the engagement would evolve?
Did you correctly perceive the positions and geometries of the enemies relevant to your current task?	Did you correctly comprehend the flight's tactics and game plan?	Did you correctly anticipate the actions, roles and duties of your flight members?
Did you correctly perceive the declaration and type of the enemies relevant to your current task?	Did you correctly comprehend if the flight was following the directed tactics/game plan?	Did you correctly anticipate the tactics and the game plan related to ranges and other decision points?
Did you correctly perceive which enemies relevant to your current task were targeted and non-targeted?	Did you correctly comprehend how the enemy presentation evolved during the engagement?	Were you able to generate alternative courses of actions or tactical modifications against possible enemy presentation changes?
Did you correctly perceive your flight members' and your own search/targeting task?	Did you correctly comprehend your flight members' and your own tactical status?	Did the final outcome of the engagement match the outcome you had anticipated?

After each trial, the flying mission was reconstructed. The reconstruction included not just the participant's cockpit displays and audio, but also the ground truth view of all simulation entities. The participant and an instructor pilot monitored the reconstruction. Once the instructor identified an event or activity associated with an individual concept, the reconstruction was paused. While stopped, the relevant RVP probes were introduced, and the instructor initiated an interview to externalize the participant's knowledge about the concept in question. During the interview, the participant – assisted by the flight instructor – compared his knowledge about the concept with the ground truth. Once the difference between the participant's knowledge and the ground truth was assessed, the instructor determined the RVP score. The procedure was repeated until the whole mission was reviewed and all RVP and RVP level scores were obtained.

Results

Data were analyzed using IBM™ SPSS™ software (version 24). For the whole flying mission, the mean RVP score was 38.18 (SD=4.65). The mean RVP scores were 13.25 (SD=1.46) for engagement 1; 13.00 (SD=2.06) for engagement 2; and 12.00 (SD=2.78) for engagement 3. The mean RVP level scores for the full mission were 13.93 (SD=1.72) for level 1; 13.46 (SD=1.64) for level 2; and 10.79 (SD=2.44) for level 3. Table 2 summarizes the descriptive statistics of the RVP level scores for each engagement.

Table 2. Minimums (Min), maximums (Max), means (M), and standard deviations (SD) of RVP level scores for each engagement (N=28).

	SMs level 1				SMs level 2				SMs level 3			
	Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD
Engagement 1	3	5	4.79	0.6	4	5	4.86	0.4	3	5	4.21	0.6
Engagement 2	3	5	4.57	0.8	3	5	4.54	0.7	0	5	3.43	1.4
Engagement 3	2	5	4.57	0.8	1	5	4.07	1.2	0	5	3.14	1.5

Only RVP level 3 scores for the full mission met the assumption of normality, hence non-parametric statistics were used to analyze the data. Based on a Friedman test, there were significant differences between the RVP scores across the engagements ($\chi2(2)=9.283$, $p=0.010$, $W=0.166$). A post-hoc Wilcoxon signed-rank test revealed significant differences for the RVP scores between engagements 1 and 2 ($z=-2.556$, $p=0.011$) and engagements 1 and 3 ($z=-3.239$, $p=0.001$).

Table 3 summarizes the differences for RVP level scores between the engagements. Across the engagements, there were significant differences between RVP level 2 scores ($\chi2(2)=9.660$, $p=0.008$, $W=0.173$) and RVP level 3 scores ($\chi2(2)=7.279$, $p=0.026$, $W=0.130$), but not between RVP level 1 scores ($\chi2(2)=2.579$, $p=0.275$, $W=0.046$).

Table 3. Pairwise comparisons between RVP level scores across engagements (N=28).

	SM level 1		SM level 2		SM level 3	
	z	*P*-value	z	*P*-value	z	*P*-value
Engagement 2 - Engagement 1	-1.561	0.119	-2.081	0.037	-2.375	0.018
Engagement 3 - Engagement 1	-1.303	0.193	-2.914	0.004	-2.869	0.004
Engagement 3 - Engagement 2	0.000	1.000	-1.939	0.053	-1.036	0.300

Discussion

For the pilots to be effective in air combat, they need to understand the components of air combat system, how these components function and how the system behaves over time (Paddon, 1977). To achieve this, their SMs (Endsley, 2000) should be closely aligned with objective reality. In this study, the RVP technique (Willis et al., 1991) was used to evaluate the participants' SMs relevant to their air combat environment, and the appropriate activities and behaviors within that environment (Wilson & Rutherford, 1989). The evaluation was conducted by measuring the pilot's knowledge about the relevant concepts and by comparing their knowledge with the ground. The scores from the evaluation results obtained with the RVP technique were used to evaluate the level to which the participants' SMs and the objective air combat reality were aligned.

The RVP scores obtained in the study consistently reflected the manipulation of the engagements' complexity. Significant differences were found between the RVP scores for engagements 1 and 2, and between engagements 1 and 3. Although the order of the engagements was not randomized and some participants had flown training missions prior to their trial, it is unlikely that the participants' cognitive strain significantly impacted the results; the duration of the mission was just over seven minutes and compared to a normal training flight, the cognitive demand of the trial mission was modest.

The RVP technique was sensitive to the task complexity manipulation when the RVP level scores were analyzed (see Table 2). It was found that the RVP technique was able to differentiate SM levels 2 and 3 between engagements 1 and 2 and between engagements 1 and 3 (see Table 3). These findings highlight the importance of formulating the probes such they unambiguously tap the pilots' knowledge about concepts associated with different SM levels. Overall, the RVP technique proved to be a useful technique to evaluate the accuracy and correctness of fighter pilots' SMs in a natural task environment.

There are some limitations, however, when using the RVP technique. If the accuracy and correctness of SMs are used as a selection or rating criterion for career progression, a pilot may be tempted to provide false information about his/her knowledge. Also, it requires a skilled flight instructor to conduct the interview procedure and to elicit the knowledge possessed by the pilot. Despite these limitations, interview techniques such as RVP are invaluable: the pilots' knowledge may be difficult to identify with performance observations, as knowledge and understanding are not necessarily manifested by overt behavior. Finally, the RVP technique is unlikely to capture the full spectrum of SMs and the content of a pilots' knowledge about the concepts. Due to these restrictions, it is likely that the RVP technique is limited to non-punitive settings, such as testing and evaluation. Even then, the number of probes used must be carefully balanced between the time required for the interview and the inclusiveness of the probes used.

Nevertheless, the RVP technique is a suitable technique to measure pilots' knowledge about the concepts in a fast-paced operational setting. It is acknowledged that the sensitivity of the RVP technique is dependent on the pilots' ability to recall past events. However, even novice fighter pilots have been repeatedly exposed to mission reconstructions where a similar post-trial recall methodology is used. After all, the pilots' ability to recall past events, supported by their notes, cockpit recordings and the access to the ground truth, forms the very foundation of any fighter pilot training (Waag & Houck, 1994).

The probes were tailored for the air combat mission used in this study. The SMs of interest dictate the concepts and the content of the probes. If the objective is to evaluate the pilots' SMs with different tactics or in different scenarios, the probes must be selected such that they capture the relevant knowledge about the concepts related to every scenario or tactics of interest. Overall, RVP is a practical technique to measure pilots' knowledge about the concepts in a virtual air combat simulation, while avoiding most of the typical challenges related to the freeze-probe, observer and self-rating techniques (Salmon et al., 2009). Moreover, the RVP technique may be used to measure knowledge about the concepts in any domain where the real-time activity cannot be paused for the time the data are collected, the activity can be reviewed ex post facto and the ground truth is available to assist in the evaluation of accuracy and correctness of SMs.

While this study investigated the SMs of individual pilots, they typically operate as a team. The future research should investigate if the RVP technique could be used to measure the similarity and accuracy of team SMs as well (Converse et al., 1991; Langan-Fox et al., 2000; Salas et al., 1994) as well.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors have no financial interest or benefit arising from applications of this research. Major Jaakko Salomäki assisted in collecting the experimental data.

References

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social forces,* 70(3), 601-636.

Converse, S., Cannon-Bowers, J., & Salas, E. (1991). Team member shared mental models: A theory and some methodological issues. In *Proceedings of the Human Factors Society Annual Meeting*, 35(19), 1417-1421.

Endsley, M. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *The International Journal of Aviation Psychology,* 3(2), 157-168.

Endsley. M. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors,* 37(1), 32-64.

Endsley, M. (1995b). Measurement of situation awareness in dynamic systems. *Human factors*, 37(1), 65-84.

Endsley, M. (2000). Situation models: An avenue to the modeling of mental models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(1), 61-64.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological review, 87*(3), 215-251.

Geiselman, R., Fisher, R., MacKinnon, D., & Holland, H. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology,* 99(3), 385-401.

Gilbert, S. (2011). *Models-based science teaching: Understanding and using mental models*. NSTA press, Arlington, VA.

Houck, M., Whitaker, L., & Kendall, R. (1992). A cognitive classification of pilot performance in air combat. In *Proceedings of the IEEE 1992 National Aerospace and Electronics Conference NAECON*, 503-509.

Jones, N., Ross, H., Lynam, T., & Perez, P. (2014). Eliciting mental models: a comparison of interview procedures in the context of natural resource management. *Ecology and Society*, 19(1), 1-7.

Kraiger, K., & Wenzel, L. (1997). Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In Brannick, M., Salas, E., & Prince, C. (Eds.) *Team performance assessment and measurement,* Lawrence Erlbaum, Mahwah, NJ, 63-82.

Langan-Fox, J. (2002). Communication in organizations: Speed, diversity, networks, and influence on organizational effectiveness, human health, and relationships. *Industrial, Work & Organizational Psychology*, 188-205.

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Techniques, methods, and analytic approaches. *Human factors,* 42(2), 242-271.

Langan-Fox, J., Anglim, J., & Wilson, J. (2004). Mental models, team mental models, and performance: Process, development, and future directions. *Human Factors and Ergonomics in Manufacturing & Service Industries,* 14(4), 331-352.

Mansikka, H., Virtanen, K., & Harris, D. (2019). Dissociation Between Mental Workload, Performance, and Task Awareness in Pilots of High Performance Aircraft. *IEEE Transactions on Human-Machine Systems,* 49(1), 1-9.

Moray, N. (1998). Identifying mental models of complex human–machine systems. *International Journal of Industrial Ergonomics,* 22(4-5), 293-297.

Morgan, M., Fischhoff, B., Bostrom, A., & Atman, C. (2002). *Risk communication: A mental models approach*, Cambridge University Press, Cambridge.

Paddon, H. (1977). *Maneuvering target simulation for testing the terminal guidance of air-to-air missiles*. Masters thesis. Air Force Institute of Technology, Columbus, OH. Available online: https://apps.dtic.mil/dtic/tr/fulltext/u2/a039757.pdf.

Pritchett, A., & Hansman, R. (2000). Use of testable responses for performance-based measurement of situation awareness. In, Endsley, M. & Garland, D. (Eds.) *Situation awareness analysis and measurement*, Lawrence Erlbaum Associates, Mahwah, NJ, 189-209.

Royal Koreal Air Force (RoKAF). (2005). *Basic Employment Manual F-16C Koran Air Force Tactics, Techniques and Procedures 3-3,* Korean Air Force.

Salas, E., Stout, R., & Cannon-Bowers, J. (1994). The role of shared mental models in developing shared situational awareness. In, *Proceedings of a CAHFA conference*, 297-304.

Salmon, P., Stanton, N., Walker, G., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring Situation Awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics,* 39(3), 490-500.

Stout, R., Cannon-Bowers, J., Salas, E., & Milanovich, D. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human factors,* 41(1), 61-71.

Sweeney, L., & Sterman, J. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, 16(4), 249-286.

Waag, W., & Houck, M. (1994). Tools for assessing situational awareness in an operational fighter environment. *Aviation, Space, and Environmental Medicine*, 65(5), A13-A19.
Wickens, C., Lee, J., Liu, Y., & Becker, S. (2004). *An introduction to human factors engineering*, Pearson Education, Upper Saddle River, NJ.
Willis, G., Royston, P., & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, 5(3), 251-267.
Wilson, J., & Rutherford, A. (1989). Mental models: Theory and application in human factors. *Human Factors*, 31(6), 617-634.

Contact Information

Dr. Heikki Mansikka (heikki.mansikka@aalto.fi)