Dissociation Between Mental Workload, Performance, and Task Awareness in Pilots of High Performance Aircraft

Mansikka, H. P., Virtanen, K. & Harris, D.

Original citation & hyperlink:
Mansikka, HP, 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 , vol. 49, no. 1, pp. 1-9.
https://dx.doi.org/10.1109/THMS.2018.2874186

DOI 10.1109/THMS.2018.2874186
ISSN 2168-2305
ESSN 2168-2291 [1]

Publisher: Institute of Electrical and Electronics Engineers (IEEE)

The Dissociation Between Mental Workload, Performance and Task Awareness in Pilots of High Performance Aircraft

H. Mansikka, K. Virtanen, and D. Harris

***Abstract*— The purpose of this study was to demonstrate how the inclusion of a tactical task goal awareness measure complemented mental workload and performance measures in a simulated air combat mission. It was hypothesized that the evaluation of the tactical task goal awareness could provide additional information concerning the cognitive demands a pilot is exposed to during a complex air combat task. A ~~novel~~ test setting was developed to test this hypothesis in a virtual flight training device. To highlight the impact of task complexity, high performance aircraft pilots' heart rate (HR), inter-beat-interval (IBI) and performance in two simple flying tasks were first compared. Then, a similar comparison, complemented with the tactical task goal awareness measure, was made with two complex flying tasks. It was found that when the pilot's awareness of the tactical goals was low, a combination of low performance and low mental workload occurred. It was concluded that when the pilots' performance is evaluated on a complex air combat task, the awareness of the tactical goals, performance and mental workload should be studied together as the pilot's awareness can explain some of HR/IBI responses that could be otherwise misinterpreted. More generally, mental workload, performance and task goal awareness should all be considered when the operator's performance in any complex human-machine system is assessed.**

***Index Terms*— air combat, aerospace simulation, mental workload, system performance, task goal awareness.**

I. INTRODUCTION

The association between mental workload (MWL) and performance has gained a great deal of interest among human factors scholars when operator's performance is assessed in a human-machine system [1]-[9]. In this study, a high performance aircraft pilot is considered as the operator in an air combat, which in turn forms a complex human-machine system. The basic motivation for the investigation of this association in a flying context is related to the potentially catastrophic effects that an excessive task load during a flight can have on the pilot performance, safety and mission effectiveness [10]-[13]. Variations in a pilot's task load tax limited cognitive resources [14], [15]. That is, each time a pilot's mental resources are expended, it is done at a certain cost [16]. The level of effort and the proportion of resources invested to meet both the objective and subjective performance criteria – mediated by the task demands – constitutes MWL [11], [17], [18]. If performance is not limited by insufficient information, a pilot's performance will eventually degrade when the resource demands of the task exceed the mental resources invested on a task [14], [19]-[22]. Once the pilot has no more excess cognitive capacity left, any increase in task demand will gradually degrade the pilots' performance – regardless of their willingness to invest more effort on the task [1], [11], [13]. In addition, the capacity and quality of cognitive resources are not the same in all people; while delivering similar levels of performance, people can have different amounts of spare cognitive capacity remaining, i.e., one cannot be defined by measuring the other.

A variety of methods to measure MWL have been developed. These include behavioral, subjective and objective measures, each with their own strengths and weaknesses. For a review, see, e.g., [23]. While MWL cannot be measured directly, variations in arousal, effort and general activation level create physiological changes which can be used as indexes of mental workload [24]-[28]. An increased heart rate (HR) and decreased inter-beat-interval (IBI) are among the measures often used as indirect indicators of an increased MWL [16], [29] and have been successfully utilized in a flying context on many occasions [7], [30]-[41]. HR can be measured by detecting heart beats from the electrocardiogram (ECG). Once HR has been determined, IBI can be analyzed by measuring the time intervals between the successive heart beats (typically the intervals between the peaks of the R wave). In empirical studies of pilot MWL in a highly demanding applied setting, HR/IBI has been found to be more sensitive to variations in task demand than some more sophisticated measures, such as the square root of the mean squared differences between successive normal-to-normal (NN) heart beat intervals, the number of successive NN interval pairs that differ by more than 50ms (or that number divided by the total number of NN intervals), and the normalized low and high frequency components of heart rate variability [7], [42].

When a pilot's performance in a complex flying task, or an operator's performance in any complex human-machine system, is evaluated, the association between MWL and performance should be approached with caution; any combination of high or low levels of MWL, high or low awareness of the tactical task goals and high or low level of performance is possible [43]-[48]. Depending on the pilots' level of awareness of the tactical task goals they may, or may not, be capable of selecting the most appropriate responses [49]. For example, pilots who are unaware of the increased requirements of their task may see no reason to invest more compensatory, voluntary cognitive effort to improve their performance [16], [17], [50]-[53]. This can lead to a situation characterized by low actual awareness of the task goals but high perceived awareness, hence low MWL and low performance. For additional information regarding the concept of 'unknown unknowns', see [54]-[58]. As a result, an evaluation of only a fighter pilot's performance and MWL can provide an incomplete picture of the pilot's mental load and the safety margin of the flight.

For a fighter pilot to be able to make correct decisions and to perform effectively, s/he needs to have sufficient understanding of the tactical situation and the associated task goals [49], [59], [60]. Awareness of the tactical task goals can be viewed as a pilot's cognitive representation of task goal-related information. As such, it is closely related to concepts like 'awareness of the critical mission requirements' and 'acquirement, assimilation and interpretation of (mission) critical information' [61], [62]. The awareness of the tactical task goals directs the pilot's subsequent sampling of information from the tactical environment, including attention allocation, decision making and task goal oriented responses [63], [64]. Based on this cyclical interaction between the pilot and the tactical environment, the pilot's responses to discrete events can be considered as manifestations of the pilot's awareness of the tactical task goals. In other words, it is possible to identify pilot responses which can only be completed with a sufficient awareness of the tactical task goals. When such discrete responses are identified, they can be used as an index of pilot's awareness of the tactical task goals. These observable pilot actions are called testable responses. For the testable responses to be effective, they should be chosen to capture the essential actions related to the activity of interest, i.e., the observable pilot responses should be dependent on the sufficient awareness of the tactical task goals [28], [65]-[69].

The aim of this study was to demonstrate how tactical goal awareness can complement MWL and performance measures in a complex air combat task. It was hypothesized that an increase in task demand during a relatively simple flying task would replicate the findings of earlier studies; MWL increases while the performance eventually decreases. In comparison, it was hypothesized that such an association between MWL and performance could not be directly transferred to a complex air combat task. It was expected that the evaluation of tactical ~~task~~ goal awareness could provide additional information concerning the cognitive demands a pilot is exposed to during the complex air combat task.

II. Method

A. ***Participants***

Thirty-seven combat ready Finnish Air Force F/A-18C/D pilots volunteered to take part. All participants were male and their average age was 31.9 years (Min=27, Max=40, SD=2.9). The participants' background varied from a wingman to a weapons instructor. Participants' mean experience with the F/A-18C/D was 686 flight hours (SD=329). Each participant had passed an extensive aeronautical medical examination within the last 12 months and they were fit to fly at the time of the study. A written, informed consent was collected from each participant. The consent was formulated in accordance with the guidance of the Finnish National Board on Research Integrity. The study was submitted for ethical approval on 6 May 2014 and was approved on 29 May 2014. A structured proforma was used to collect background data and information concerning participants' relevant activities for the 12 hours prior to participating. Data from five participants were lost due to corrupted ECG samples and flight training device malfunctions. As a result, the analysis was based on data from 32 participants.

B. ***Test Design***

An F/A-18C/D Weapon Tactics and Situation Awareness Trainer (WTSAT) was used for a flying task. WTSAT is a non-motion virtual flight training device with a 135-degree visual and fully functional cockpit and avionics. While the lack of moving-base simulation does not allow WTSAT to be classified as a flight simulator per se, it replicates the F/A-18C/D flying characteristics with such a high accuracy that Finnish Air Force F/A-18C/D pilots can use it to fly basic and advanced training missions as well as instrument and combat check rides. Figure 1 shows WTSAT in a typical training configuration.

FIGURE 1

WTSAT IN A TYPICAL TRAINING CONFIGURATION

The flying task was a typical defensive counter air mission which started from ground and progressed as a continuous, non-stop combat mission consisting of take-off, ingress, air combat, egress and instrument approach segments. The first three segments of the mission were flown as a flight. A ‘flight’ in standard Air Force usage refers to a unit of four aircraft, which is composed of two ‘elements’ with two aircraft in each. Each participant was designated as the wingman of the second element.

The egress and instrument approach segments included both two-ship and single-ship operations. To standardize the mission flow between the participants, all other aircraft in the simulation were constructive simulation entities and followed a predefined script; the enemy aircraft followed predefined trajectories whereas the friendly aircraft maneuvered in accordance with established tactical procedures. The radio transmissions were prepared as an audio file which was time-synchronized with the flight trajectories and activities of the constructive simulation entities. The simulated radio communications included transmissions from three different air traffic controllers (tower and two radar controllers), two different fighter controllers and the transmissions from simulated members of the friendly flight, i.e., the leader of the lead element, the wingman of the lead element and the leader of the second element. All simulated communications were recorded using the real tower, radar, fighter controller and aircraft radios. To increase the sense of authenticity, the radios were operated by real air traffic controllers, fighter controllers and F/A-18C/D pilots. In addition to the normal radio traffic, the audio file included radio jamming and radio noise. The participants were given directive and informative calls on the radio and they were expected to reply to them. The participants’ missile shots had no effect on the threat aircraft. Instead, the threat aircraft were removed from the aerial picture based on the predefined script. The participants were not aware of their missiles’ probability of kill being set to zero as this might have affected their willingness to invest effort on the task. Because of the scripted test setting, the mission evolved similarly for each participant and the only variations to the mission complexity, and to the flight’s performance resulted from the participants’ own actions.

Before the trial the participants had 15 minutes to study the mission briefing material, which was provided as hard copy. To provide each participant with equal detail of the mission, no clarifications were given, and no questions were answered during the briefing. After the participants had studied the briefing material, the simulation was activated. The participants did not receive any kind of mission specific training prior to the trial because the mission was planned to be used as a check ride where the participants' proficiency in their current duty was to be evaluated. The mission consisted of tasks that the participants were expected to be familiar with, and any mission specific training would have destroyed the inherent nature and purpose of the check ride. Also, as the mission was designed to comprise a whole, logical air combat mission, any variations in the sequence of the events would have compromised the tactical and spatial immersion. The mission duration from the engine start-up to final landing was approximately 45 minutes.

The study had a within-subject design. Participants' HR/IBI and performance data were retrieved from two instrument landing system approach segments (or 'Approaches') and from two beyond-visual-range (BVR) attack segments (or 'Commits'). There were low difficulty (LO) and high difficulty (HI) Approach segments and Commit segments. In addition, the participants' awareness of the tactical task goals during the Commits was recorded and used as testable responses. The Approaches represented simple, routine flying tasks, whereas the Commits represented complex, highly dynamic tasks. The task demands of the Approaches and Commits was used as an independent variable. Participants' HR/IBI values, performance scores and testable response values were used as dependent variables.

C. ***Independent Variable Manipulation***

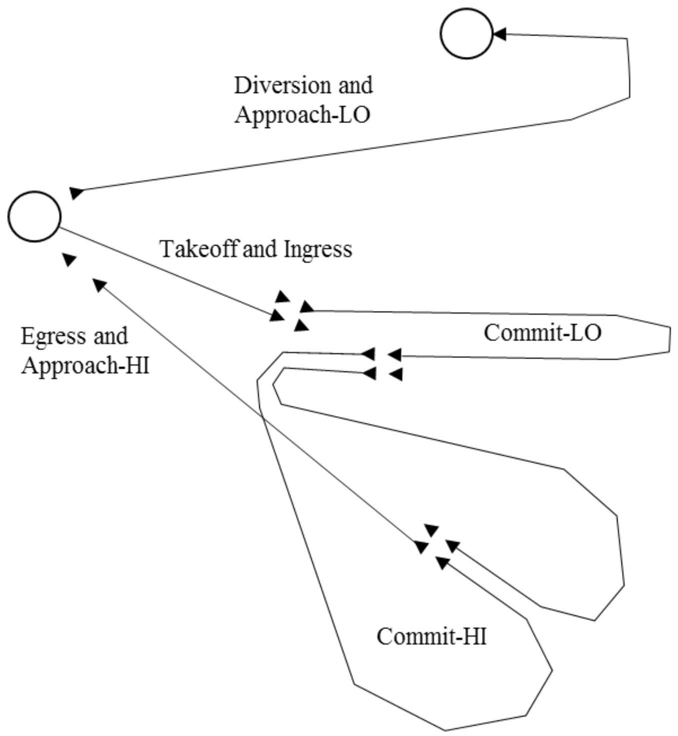
Participants flew two Approach and two Commit segments, both with HI and LO scenarios. The HI Approach (Approach- HI) was designed to have a higher task demand than the LO Approach (Approach-LO). This was achieved by manipulating the temporal demand just before the participants commenced the approach, and the distracting factors before and during the approach. Approach-HI was flown to a planned landing field immediately after the mission's combat phase and high-speed egress. Approach-1 was flown as a wingman in a two-ship radar trail. This meant that while flying the normal approach profile, the participants had to maintain the radar track of the leading aircraft and maintain a predefined separation to it. The lead aircraft in the two-ship radar trail reduced its airspeed from tactical speed to approach speed at the last possible moment, thus limiting the time the participants had available for approach preparations. During Approach-HI, the participants were also required to respond to tactical radio calls. In summary, these additional tasks and the increased temporal demand made the task demands of Approach-HI higher than those of a typical approach with a similar profile. While the individual tasks during Approach-HI were simple, together they had the potential to seriously compromise the participants' performance if their attention was not divided properly.

In each trial, Approach-LO was flown to an alternate airfield following a diversion decision after Approach-HI. The lower task demand of Approach-LO was achieved by allowing the participants to fly it as a single-ship, letting them choose their airspeed and flight profile before starting the actual approach and excluding any external distractions during the approach. Both Approach-HI and Approach-LO had almost identical, standard approach profiles. Both approaches were flown in instrument meteorological conditions (IMC) with 0 m (0 ft) IMC visibility. For Approach-HI, the cloud base was set below the landing minimum, hence forcing a go-around. For Approach-LO, the cloud base and runway visibility were set to allow a full stop landing.

Both Commits (Commit-LO and Commit-HI) were flown as a flight with the participants flying as the wingman of the second element. While Commit-LO and Commit-HI were both BVR attacks, they had different threat presentations and thus required different actions from the friendly four-ship. Commit-LO was designed to have a lower task demand, to be less dynamic and complex and to have lower demands for the tactical task goal awareness than Commit-HI. This was achieved by manipulating the scripted enemy formation and maneuvers, the flow of the friendly aircraft, timing of radio transmissions and the quality of available sensor information. During the pre-testing, it was verified that in both Commits there was enough information and time available to achieve a satisfactory outcome. The duration of each Commit was approximately five minutes. Both Commits were flown in full IMC. Figure 2 summarizes the flow of the flying task. The triangles depict the friendly aircraft and the solid lines represent their general flight path. The circles represent the planned and alternate airfields.

FIGURE 2

SUMMARY OF THE FLOW OF THE FLYING TASK WITH DIFFERENT SEGMENTS



D. ***Dependent Variable Measures***

ECG recording, analysis and reporting were conducted in accordance with the recommendations by the Task Force of The European Society of Cardiology and the North American Society of Pacing and Electrophysiology [70]. Mind Media Nexus-10 MKII system was used for the ECG recording. Three electrodes were placed below the left (negative) and right (ground) clavicle and the left costal cartilage (positive). Data were first recorded using Biotrace+ software (version V2012C) from where the samples were exported to Kubios HRV 2.2 software for artefact removal and further analysis. As recommended by [70], the artefacts were manually identified and removed from the ECG samples. The artefacts in the ECG may cause significant errors to analysis, especially when short (two to five minute) ECG samples are used. Baseline wander, atrial fibrillation, ectopic beats, irregular heart rate and multiplied or masked heart beats, erroneous detections of heart beats and line interference are among the typical artefacts [71]. For each participant, a five-minute average of HR/IBI were recorded and used for the analysis. A sampling rate of 1024 Hz was used for all samples.

The participants' discernible and identifiable reactions to their targeting responsibilities were used as the testable responses for the measuring of the awareness of tactical task goals. Targeting is a directive task to locate a specified enemy group from the aerial picture, to allocate on-board sensors and weapons against it and to eventually commit against the group. The following pilot reactions were used as testable responses: adjusting the on-board sensors to search the correct area in space, identifying the correct target from the aerial picture, creating an attack geometry against the correct target and employing weapons against it. The targeting command was provided to the participants as a radio call within the audio file. In addition to the radio call, the participants could determine their targeting task by relying on a number of pre-determined decision criteria, e.g., the flow and targeting of other friendly aircraft, enemy formation and maneuvers, and vertical/horizontal separation between friendly and enemy aircraft. As a result, the testable responses were an aggregate of various aspects of tactical goal awareness.

A total of four targeting tasks were used for the measuring of the awareness of tactical task goals. Each Commit included two targeting tasks used as testable responses. That is, the pilot's overt responses to discrete targeting tasks were used as manifestations of the pilot's awareness of these tactical task goals. The pilots' overall awareness of the tactical task goals in each Commit was calculated as a sum of the individual targeting scores. For each testable response, the participants' actions were rated either as correct (value of 1) or wrong (value of 0). The testable responses were communicated as a percentage value of the maximum score. The responses were scored by a subject matter expert (a qualified F/A-18C weapons instructor) as suggested by [72].

The performance scores of Commits were based upon the participants' reactions and the mission's outcome. As long as the agreed tactical procedures were followed and the mission's outcome met the mission objectives, the participants were free to use the maneuvers they considered most appropriate and they had the liberty to use the sensors and weapons as they considered proper. Therefore, only the items that were tactically mandatory, were clearly observable and that could be unambiguously defined as correct or incorrect were scored. Tactical reactions were scored based on whether they were conducted correctly and safely. Unsafe, incorrect or missed tactical reactions were given a score of '0' whereas each safe and correct response was given a score of '1'. The participants' overall performance score in each Commit was formed by summing up the respective scores. The minimum performance score was '0' whereas the maximum performance score was '60'. The participants' performance scores in Commits were communicated as percentage values of each Commit's maximum performance score.

The Approach performance was based upon deviations from the predefined target speed along with the localizer (horizontal) and glideslope (vertical) errors from the optimal approach trajectory. Each error component was scored independently and their values ranged from '5' (best performance) to '0' (worst performance). The deviations from the target values were recorded and scored every 0.5 nm (0.9 km) between 5.5 nm (10.1 km) and 0.5 nm (0.9 km) from the runway threshold. Participants' Approach performance scores were generated by calculating the mean of the performance scores from each data collection point. Both Approaches were scored using an official instrument check ride rating scale. The scores were communicated as percentage values of each Approach's maximum performance score.

III. Results

Data were analyzed using IBM SPSS Statistics software (version 22). Data values more than 1.5 interquartile ranges below the first quartile or above the third quartile were excluded from the analysis. Normality of the distributions of the data was verified using the Shapiro-Wilk test.

A repeated-measures nested ANOVA with Approach/Commit and high/low difficulty as main effects was performed on the performance data. There was a significant overall difference between performance in the Approach and Commit segments ($F_{1,31}=154.053$, $p<0.001$). Overall, Approaches (M=75.6) elicited superior performance to Commits (M=50.9). There was also a main effect of difficulty ($F_{1,31}=62.378$, $p<0.001$). As may be expected, LO segments (M=69.6) produced better performance than HI segments (M=56.9). There was no significant interaction term ($F_{1,31}=0.021$, $p=0.886$).

The performance data were further analyzed with paired t-tests. Table 1 presents the descriptive statistics of the performance scores. The pairwise comparisons of performance means indicated a significant difference between Approach-HI (M=69.1, SD=17.4) and Approach-LO (M=82.1, SD=8.3); $t(31)=-3.886$, $p<0.001$. The minimum performance on Approach-HI was 4.0%, as one participant almost completely missed the approach profile after being severely distracted by the additional tasks. A significant difference in performance was also found between Commit-LO (M=57.0, SD=10.4) and Commit-HI (M=44.8, SD=8.0); $t(31)=5.561$, $p<0.01$. All pairwise comparisons are summarized in Table 4.

TABLE I

MINIMUMS (MIN), MAXIMUMS (MAX), MEANS (M) AND STANDARD DEVIATIONS (SD) OF PERFORMANCE SCORES IN DIFFERENT MISSION SEGMENTS (N=32).

	Performance (%)			
	Min	Max	M	SD
Commit-LO	40.0	75.0	57.0	10.4
Commit-HI	32.0	68.0	44.8	8.0
Approach-HI	4.0	87.1	69.1	17.4
Approach-LO	62.4	93.9	82.1	8.3

Table 2 presents the descriptive statistics for the IBI and HR values. Further repeated-measures nested ANOVAs with Approach/Commit and high/low difficulty as main effects was performed on the workload data. There was no significant overall difference in HR between the Approach and Commit segments ($F_{1,31}$=2.760, p=0.107). Mean HR value for Approach segments was 95.5; mean HR value for Commit segments was 93.6. There was also no overall significant difference in HR values between HI and LO segments ($F_{1,31}$=2.275, p=0.142). Mean HR values for LO segments was 95.1; mean HR values for HI segments was=94.0. There was, however, a highly significant interaction term ($F_{1,31}$=32.316, p<0.001) – see Table 2 for mean values. Further decomposition of the interaction term using post-hoc t-tests showed that for mean HR values, there was a significant difference between Commit-LO and Commit-HI; t(31)=7.608, p<0.001.

There was no significant overall difference in IBI values between the Approach and Commit segments ($F_{1,31}$=2.583, p=0.118). Mean IBI value for Approach segments was 651.7ms; mean IBI value for Commit segments was 665.0ms. There was also no overall significant difference in IBI values between HI and LO segments ($F1,31=2.624$, $p=0.114$). Mean IBI value for LO segments was 654.6ms; mean IBI value for HI segments was 662.7ms. The interaction term was significant ($F_{1,31}=40.653$, $p<0.001$) – see Table 2 for mean values. Further decomposition of the interaction term using post-hoc t-tests showed that there was a significant IBI value difference between Approach-HI and Approach-LO; $t(31)=-2.099$, $p<0.05$. A significant IBI value difference was also found between Commit-LO and Commit-HI; $t(31)=-7.356$, $p<0.001$.

The descriptive statistics of the testable responses are given in Table 3. A pairwise comparison revealed a significant difference in the means of correct testable responses between Commit-LO ($M=43.8$, $SD=50.4$) and Commit-HI ($M=12.5$, $SD=33.6$); $t(31)=2.743$, $p<0.05$.

TABLE II

MINIMUMS (MIN), MAXIMUMS (MAX), MEANS (M) AND STANDARD DEVIATIONS (SD) OF HR AND IBI VALUES IN DIFFERENT MISSION SEGMENTS (N=32).

	MEAN HR (1/min) / MEAN IBI (ms)			
	Min	Max	M	SD
Commit-LO	72.7/435.2	138.1/830.8	95.7/649.8	17.5/110.7
Commit-HI	69.9/460.2	130.7/864.2	91.5/680.2	17.3/118.5
Approach-HI	76.7/446.4	134.6/787.1	96.5/644.1	17.8/108.0
Approach-LO	71.4/448.2	134.4/848.1	94.5/659.3	17.3/112.0

TABLE III

MINIMUMS (MIN), MAXIMUMS (MAX), MEANS (M) AND STANDARD DEVIATIONS (SD) OF TESTABLE RESPONSES IN DIFFERENT MISSION SEGMENTS (N=32).

	Testable responses (%)			
	Min	Max	M	SD
Commit-LO	0.0	100.0	43.8	50.4
Commit-HI	0.0	100.0	12.5	33.6

TABLE IV

SUMMARY OF TEST STATISTICS AND CHANGES IN PAIRWISE COMPARISONS BETWEEN MISSION SEGMENTS. ***P<0.001; **P<0.01; *P<0.05 (N=32).

	Commit-LO - Commit-HI			Approach-HI - Approach-LO		
	M	SE	t	M	SE	t
Performance (%)	-12.3	2.2	5.561**	-13.0	3.3	-3.886***
Mean HR (1/min)	4.1	0.5	7.608***	2.1	1.1	1.834
Mean IBI (ms)	-30	4.1	-7.356***	-15	7.2	-2.099*
Testable Responses (%)	31	11	2.743*	-	-	-

A Pearson product-moment correlation was run to determine correlations between HR, IBI, performance and testable responses both in Commit-LO and Commit-HI. Only HR and IBI were correlated in Commit-LO (r=-0.984, p<0.01) and in Commit-HI (r=-0.985, p<0.01). Similarly, correlations were determined between HR, IBI and performance in Approach-HI and Approach-LO. HR and IBI were significantly correlated in Approach-HI (r=-0.989, p<0.01) and in Approach-LO (r=-0.987, p<0.01).

IV. Discussion

In this study, the participants' HR/IBI and performance were measured during two Approaches and two Commits included in a simulated air combat mission. The participants obtained significantly higher performance scores in Approach-LO (the easier approach) compared to those in Approach-HI (see Tables 1 and 4). There was no significant difference in MWL (as assessed by IBI values) between the main effects of Approach/Commit or HI/LO segments. However, there was a highly significant interaction term. Following the de-composition of the interaction term, it can be seen that as the IBI values in Approach-HI were significantly lower than in Approach-LO, it was concluded that Approach-HI had a significantly higher task demand than Approach-LO (see Tables 2 and 4). The tactical, high speed radar trail egress preceding Approach-HI increased the temporal demand during the Approach preparations, increased the amount of distracting information and generated potential for an incorrect attentional focus during the initial approach. In addition, the radar trail approach itself increased the participants' task loading; the on-board radar had to be adjusted to maintain a radar track of the preceding aircraft and the separation to the lead aircraft had to be kept within acceptable limits. These factors were enough to generate a significantly higher task demand between otherwise similar approaches, as was reflected in the workload measures.

Despite these distracting elements during the initial phase of Approach-HI, the participants were highly familiar with the task and had flown the same approach numerous times both in a virtual flight training device and in real aircraft; they knew the approach profiles and procedures, they knew what information was needed to execute the approach and they knew where the information could be acquired. It can be thus assumed that any observed performance degradations were not caused by the participants not being aware of the task requirements or tactical task goals. As a result, the IBI values and the performance scores during Approach-HI and Approach-LO followed the expected pattern and replicated the findings of earlier studies [7], [30]-[41]; an increased task demand was reflected in decreased performance and increased MWL as indicated by the lowered values of IBI (see Tables 2 and

4). A similar association between operator's performance and MWL has been reported not only in the high performance aircraft domain, but in other human-machine systems as well [73], [74].

When the participants' performance scores in Commit-LO and Commit-HI were compared, Commit-LO had a significantly higher performance score average than Commit-HI (see Tables 1 and 4). As a reminder, Commit-LO and Approach-LO were designed to have a lower task demand than Commit-HI and Approach-HI, so this result is as expected. However, the HR and IBI values of Commit-LO and Commit-HI did not follow the same pattern as they did during the Approaches. Following the decomposition of the interaction term, it can be seen that the IBI values during Commit-HI were significantly higher (and the HR values were significantly lower) than during Commit-LO (see Tables 2 and 4). In other words, Commit-HI had lower performance scores *and* lower MWL. Unlike the Approaches, the Commits were highly complex and dynamic tasks with rapidly and unexpectedly changing tactical situations. These conditions generated extremely high demands on participants' ability to collect and process relevant information, to build a coherent mental representation of the tactical situation and to select their responses accordingly. The results of this study indicated that the lower value of awareness of the tactical task goals, measured using the testable responses, was associated with lower performance and lower MWL (see Tables 1-4). It is possible that as the awareness of the tactical task goals was lower, the participants failed to allocate their attention to effectively sample the tactical environment [63], [64]. As a result, the participants may not have obtained the necessary information related to the tactical situation and may have had difficulties in conducting the higher level mental processing required for the establishment of the mental representation of the tactical situation [10], [12], [49], [51], [54], [56]. In other words, when the participants were unaware of the tactical task goals, they probably did not excessively consume their mental processing capacity and thus their MWL remained low.

If MWL were to be evaluated during day-to-day operations, measuring HR/IBI alone would probably not be appropriate. Since various physiological measures provide unique information about the various aspects of operator's MWL, any single measure is unlikely to reflect all the task demands relating to the different cognitive resources of the pilot [35]. That said, while HR and IBI are unable to distinguish the exact phase or modality of information processing that is being loaded, they provide an assessment of the overall workload [22], [23]. If a certain modality, phase or cognitive resource would be of interest, some other physiological measure would likely to be more diagnostic. Alternatively, physiological measures of MWL can be supplemented by post-task interviews or the administration of multi-dimensional subjective measures, e.g. the NASA-TLX [75] or SWAT – Subjective Workload Assessment Technique [76]. Furthermore, if it is required to measure MWL in real flight, physiological measures might become insensitive altogether as factors like extreme cockpit temperatures, exposure to direct sunlight and high G-loads can generate physiological responses which can be falsely interpreted as MWL responses. The question whether MWL should be measured using multiple physiological measures, behavioral measures or subjective measures (or any combination of them) is dependent on the test environment and operational context. As the objective of this study was to show how the tactical task goal awareness measure can complement MWL and performance measures, the use of HR/IBI was considered appropriate.

The traditional assumption is that when task demands are increased, and pilots are capable and willing to invest more mental effort, MWL gradually increases [22] and this increase in MWL is manifested by increased HR and lowered IBI [7], [16], [30]-[41]. However, as indicated by the results of this study, the association of MWL and performance is more complex. Based on the mean values for performance, IBI/HR and testable responses in Tables 1-3, and significant differences in the pairwise comparisons in the post-hoc t-tests between the mission segments (see Table 4), the following combinations of task demands, MWL, task goal awareness, and performance were found: lower task demands, lower MWL and higher performance (Approach-LO); higher task demands, higher MWL and lower performance (Approach-HI); higher task demands, lower MWL, lower awareness of the tactical task goals and lower performance (Commit-HI); lower task demands, higher MWL, higher awareness of the tactical task goals and higher performance (Commit-LO). As suggested by [43]-[48] other combinations can also exist. That is, high performance can encourage a sense of low (perceived)

MWL and high (perceived) awareness of the tactical task goals and tactical situation. Alternatively, low performance can encourage a sense of high (perceived) MWL and low (perceived) awareness. As suggested by the findings of this study, it is not possible to draw conclusions about one by measuring the other; pilots with overtly similar performance may experience significantly different levels of MWL and their awareness of the tactical task goals can also be significantly different. This conclusion is also supported by the finding that the performance data, the workload data and the testable responses did not correlate within single segments of the flying task. In fact, any combinations of these three factors are possible, and if any of the factors falls below a satisfactory threshold, mission effectiveness and flight safety may become compromised.

If just MWL and the performance are being evaluated during a complex air combat mission, or any complex human-machine system operation, a dissociation between the two may be found which is difficult to understand [77]-[80]. But when the pilot's tactical goal awareness is considered as an additional measure, the conclusion may be very different; low performance could result from the pilot working at the upper limits of his/her cognitive resources thus being unable to perceive and process mission critical information. When the fighter pilot's performance in a complex air combat task is evaluated, it is therefore necessary to measure not just MWL, awareness or performance, but all of them. Put simply, from a cognitive perspective you work as hard as you think that you need to. If you are not aware of the true demands of the task, MWL may be relatively low – but ultimately so too will be performance. MWL is a product of perceived tactical task goals, not objective requirements (of which the pilot may be unaware).

Despite the previous research efforts [6], [8], [74], [78], [80], future research is still needed to assure that our understanding of the association between operator's performance, MWL and awareness remains aligned with rapidly evolving human-machine systems. For instance, emerging 5th generation fighters are likely to expose pilots to new types of tasks and cognitive demands. As demonstrated in this study, when an operator's performance is evaluated in a complex human-machine system, it is not just the operator's performance, MWL or awareness that should be measured, but all of these factors.

Therefore, the comparison and evaluation of the performance of complex human-machine systems could be approached as a multicriteria decision analysis problem [81], [82]. Multicriteria decision analysis tools and techniques have great potential as an approach for the evaluation of future human-machine system designs or operating procedures.

V. Conclusions

A test setting was developed to investigate the high performance aircraft pilots' performance, MWL and tactical task goal awareness in a virtual flight training device. During the relatively simple approach tasks, increased task demand was associated with lower performance and increased MWL. During the complex air combat tasks, the association between performance and MWL was less straightforward; when the pilot's awareness of the tactical goals was low, a combination of low performance and low MWL occurred. When the pilots' performance is evaluated on a complex air combat task, or when any complex human-machine system performance is being evaluated, the awareness of the tactical task goals, performance and MWL should be studied together. The pilot's awareness of the tactical task goals can explain some of the HR/IBI responses that could be otherwise misinterpreted. The findings of this study could, and should, be utilized in the evaluation of any complex human-machine system. That said, future research efforts should further validate the association between operator's performance, MWL and awareness, which can be seen as multiple conflicting criteria for the measurement of human-machine performance. Thus, multicriteria decision analysis may provide new and innovative ways to support the performance evaluation of future human-machine systems.

Acknowledgment

This study was ordered by the Finnish Air Force and was conducted jointly by the Finnish Defense Research Agency, Aalto University and Coventry University.

References

[1] R. C. Williges and W. W. Wierwille, “Behavioral measures of aircrew mental workload,” *Human Factors,* vol. 21, no. 5, pp. 549-574, 1979. DOI: 10.1177/001872087902100503.

[2] B. H. Kantowitz, “Mental workload”. *Adv. Psychol.,* vol. 47, pp. 81-121, 1987. DOI: 10.1016/S0166-4115(08)62307-9.

[3] R. J. Lysaght, S. G. Hill, A. O. Dick, B. D. Plamondon, and P. M. Linton, “Operator workload: comprehensive review and evaluation of operator workload methodologies,” Army Institute for Behavioral and Social Sciences, Alexandria, VA, USA, Tech. Rep. 851, Jun. 1989.

[4] D. De Waard, *The Measurement of Drivers' Mental Workload*. Haren, The Netherlands: The Traffic Research Centre VSC, University of Groningen, 1996, pp. 53-92.

[5] H. R .Jex, “Measuring mental workload: problems, progress and promises,” in *Human Mental Workload,* P. A. Hancock, and N. Meshkati, Eds., Amsterdam, The Netherlands: North-Holland, 1988, pp. 5-36.

[6] M. Hou, G. Ho, G. Arrabito, S. Young, and S. Yin, “Effects of display mode and input method for handheld control of micro aerial vehicles for a reconnaissance mission,” *IEEE Trans. Syst., Man, Cybern. A. Syst., Humans*, vol. 43, no. 2, pp. 149-160, Mar. 2013. DOI: 10.1109/TSMC.2013.2239595.

[7] H. Mansikka, P. Simola, K. Virtanen, D. Harris, and L. Oksama, “Fighter pilots’ heart rate, heart rate variation and performance during instrument approaches,” *Ergonomics*, vol. 59, no. 10, pp. 1344-1352, Mar. 2016. DOI: 10.1080/00140139.2015.1136699.

[8] D. Pantförder, B. Vogel-Heuser, D. Gramß, and K. Schweizer, “Supporting operators in process control tasks—benefits of interactive 3-D visualization,” *IEEE Trans. Human Mach. Syst.*, vol. 46, no. 6, pp. 895-907, Dec. 2016. DOI: 10.1109/THMS.2016.2599497.

[9] J. Zhang, Z. Yin, and R. Wang, “Nonlinear dynamic classification of momentary mental workload using physiological features and NARX-model-based least-squares support

vector machines," *IEEE Trans. Human Mach. Syst.*, vol. 47, no. 4, pp. 536-549, Aug. 2017. DOI: 10.1109/THMS.2017.2700631.

[10]J. Aasman, G. Mulder, and L. J. M. Mulder, "Operator effort and the measurement of heart-rate variability," *Human Factors*, vol. 29, no. 2, pp. 161-170, Apr. 1987. DOI:10.1177/001872088702900204.

[11]F. G. W. C. Paas, and J. J. G. Van Merriënboer, "The efficiency of instructional conditions: an approach to combine mental effort and performance measures," *Human Factors*, vol. 35, no. 4, pp. 737-743, Dec. 1993. DOI: 10.1177/001872089303500412.

[12]M. R. Endsley, "Automation and situation awareness", in *Automation and Human Performance: Theory and Applications*, R. Parasumaran, and M. Mouloua, Eds. Mahwah, NJ, USA: Lawrence Erlbaum, 1996, pp. 163-181.

[13]J. B. Noel, K. W. Bauer Jr. and J. W. Lanning, "Improving pilot mental workload classification through feature exploitation and combination: a feasibility study," Comput. Oper. Res., vol. 32, no. 10, pp. 2713-2730, Oct. 2005. DOI: 10.1016/j.cor.2004.03.022.

[14]C. D. Wickens, "Processing resources and attention," in *Multiple-Task Performance*, D. Damos, Ed. London, England, Taylor and Francis, 1991, pp. 3-34.

[15]C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, pp. 159-177, Nov. 2010. DOI: 10.1080/14639220210123806.

[16]G. R. J. Hockey, "Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework," *Biol. Psychol.,* vol. 45, no. 1, pp. 73-93, 1997. DOI: 10.1016/S0301-0511(96)05223-4.

[17]G. Mulder, "The concept and measurement of mental effort," in *Energetics and Human Information Processing*, G. Hockey, Ed. Dordrecht, The Netherlands: Springer, 1986, pp. 175-198. DOI: 10.1007/978-94-009-4448-0_12.

[18]G. Camp, F. Paas, R. Rikers, and J. van Merrienboer, “Dynamic problem selection in air traffic control training: a comparison between performance, mental effort and mental efficiency,” *Comput. Hum. Behav*., vol. 17, no. 5-6, pp. 575-595, Sep.-Nov. 2001. DOI: 10.1016/S0747-5632(01)00028-0.

[19]D. A. Norman, and D. G. Bobrow, “On data-limited and resource-limited processes,” *Cognitive Psychol.,* vol. 7, no. 1, pp. 44-64, 1975. DOI: 10.1016/0010-0285(75)90004-3.

[20]L. A. Muse, S. G. Harris, and H. S. Feild, “Has the inverted-U theory of stress and job performance had a fair test?,” *Hum. Perform*., vol. 16, no. 4, pp. 349-364, Nov. 2009. DOI: 10.1207/S15327043HUP1604_2.

[21]J. Kavanagh, “Stress and performance. A review of the literature and its applicability to the military,” Rand Corp., Santa Monica, CA, USA, Tech. Rep. ADA439046, 2005.

[22]C .D. Wickens, “Multiple resources and mental workload,” *Human Factors*, vol. 50, no. 3, pp. 449-455, Jun. 2008. DOI: 10.1518/001872008X288394.

[23]R. D. O’Donnell, F. T. Eggemeier, F., “Workload assessment methodology,” in *Handbook of Perception and Human Performance,* vol. 2, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., Oxford, England: John Wiley, 1986, pp. 1-49.

[24]H. Ursin, and R. Ursin, “Physiological indicators of mental workload mental workload,” in *Mental Workload. NATO Conference Series*, vol. 8, N. Moray, Ed. Boston, MA, USA: Springer. DOI: 10.1007/978-1-4757-0884-4_21.

[25]W. W. Wierwille, “Physiological measures of aircrew mental workload,” *Human Factors,* vol. 21, no. 5, pp. 575-593, Oct. 1979. DOI: 10.1177/001872087902100504.

[26]G. F. Wilson, “An analysis of mental workload in pilots during flight using multiple psychophysiological measures,” *Int. J. Aviat. Psychol.,* vol. 12, no. 1, pp. 3-18, Nov. 2009. DOI: 10.1207/S15327108IJAP1201_2.

[27]K. Ryu, and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *Int. J. Ind. Ergonom*., vol. 35, no. 11, pp. 991-1009, Nov. 2005. DOI: 10.1016/j.ergon.2005.04.005.

[28]C. F. Rusnock, and C. D. Geiger, "Simulation-based evaluation of adaptive automation revoking strategies on cognitive workload and situation awareness," *IEEE Trans. Human Mach. Syst.*, vol. 47, no. 6, pp. 927-938, Dec. 2017. DOI: 10.1109/THMS.2016.2618004

[29]A. J. Terkelsen, H. Mølgaard, J. Hansen, O. K. Andersen, and T. S. Jensen, "Acute pain increases heart rate: differential mechanisms during rest and mental stress," *Autonomic Neuroscience,* vol. 121, no. 1–2, pp. 101-109, Aug. 2005. DOI:10.1016/j.autneu.2005.07.001.

[30]C. Sekiguchi, Y. Handa, M. Gotoh, Y. Kurihara, Y. Nagasawa, and I. Kuroda, "Frequency analysis of heart rate variability under flight conditions," *Aviat. Space Envir. Md*., vol. 50, pp. 625-634.

[31]G. F. Wilson, and F. T. Eggemeier, "Psychophysiological assessment of workload in multi-task environments," in *Multiple-Task Performance,* D. Damos Ed. London, England, Taylor and Francis, 1991, pp. 329-360.

[32]A. H. Roscoe, "Assessing pilot workload. Why measure heart rate, HRV and respiration?," *Biol. Psychol.,* vol. 34, no. 2, pp. 259-287, Nov. 1992. DOI: 10.1016/0301-0511(92)90018-P.

[33]P. G. A. M. Jorna, "Heart rate and workload variations in actual and simulated flight," *Ergonomics*, vol. 36, no. 9, pp. 1043-1054, Jul. 2010. DOI: 10.1080/00140139308967976.

[34]A. H. Roscoe, "Heart rate as a psychophysiological measure for in-flight workload assessment," *Ergonomics*, vol. 36, no. 9, pp. 1055-1062, Jul. 2010. DOI: 10.1080/00140139308967977.

[35]G. F. Wilson, "Air-to-ground training missions: a psychophysiological workload analysis," *Ergonomics*, vol. 36, no. 9, pp. 1071-1087, Jul. 2010. DOI: 10.1080/00140139308967979.

[36]J. A. Veltman, and A. W. K. Gaillard, "Physiological indices of workload in a simulated flight task," *Biol. Psychol.,* vol. 42, no. 3, pp. 323-342, Feb. 1996. DOI:10.1016/0301-0511(95)05165-1.

[37]E. Svensson, M. Angelborg-Thanderez, L. Sjöberg, and S. Olsson, "Information complexity-mental workload and performance in combat aircraft," *Ergonomics*, vol. 40, no. 3, pp. 362-380. DOI: 10.1080/001401397188206.

[38]T. C. Hankins, and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight," *Aviat. Space Envir. Md*., vol. 69, no. 4, pp. 360-367, Apr. 1998.

[39]M. Bonner, and G. F. Wilson, "Heart rate measures of flight test and evaluation," *Int. J. Aviat. Psychol.,* vol. 12, no. 1, pp. 63-77, 2002. DOI:10.1207/S15327108IJAP1201_6.

[40]S. Magnusson, "Similarities and differences in psychophysiological reactions between simulated and real air-to-ground missions," *Int. J. Aviat. Psychol.,* vol. 12, no. 1, pp. 49-61, 2002. DOI: 10.1207/S15327108IJAP1201_5.

[41]T. Lahtinen, J. Koskelo, T. Laitinen, and T. Leino, "Heart rate and performance during combat missions in a flight simulator," *Aviat. Space Envir. Md*., vol. 78, no. 4, pp. 387-391, Apr. 2007.

[42]H. Mansikka, K. Virtanen, D. Harris, and P. Simola, "Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test," *Appl. Ergon.,* vol. 56, pp. 213-219, Sep. 2016. DOI: 10.1016/j.apergo.2016.04.006.

[43]M. R. Endsley, "Situation awareness and workload - flip sides of the same coin," in *Proceedings of the Seventh International Symposium on Aviation Psychology*, R. S. Jensen, and D. Neumeister, Eds., The Ohio State University, Department of Aviation, Columbus, OH, USA, 1993, pp. 906-911.

[44]M. R. Endsley, and E. O. Kiris, "The out-of-the-loop performance problem and level of control in automation," *Human Factors*, vol. 37, no. 2, pp. 381-394, Jun. 1995. DOI: 10.1518/001872095779064555.

[45]M. R. Endsley, "Level of automation effects on performance, situation awareness and workload in a dynamic control task," *Ergonomics*, vol. 42, no. 3, pp. 462-492. DOI: 10.1080/001401399185595.

[46]D. B. Kaber, E. Onal, and M. R. Endsley, "Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload," *Hum. Factor Ergon*., vol. 10, no. 4, pp. 409-430, 2000. DOI: 10.1002/1520-6564(200023)10:4<409::AID-HFM4>3.0.CO;2-V.

[47]M. L. Cummings, and P. J. Mitchell, "Operator scheduling strategies in supervisory control of multiple UAVs," Aerosp. Sci. Technol., vol. 11, no. 4, pp. 339-348, May 2007. DOI: 10.1016/j.ast.2006.10.007.

[48]R. Sabatini, M. A. Richardson, M. Cantiello, M. Toscano, and P. Fiorini, "A novel approach to night vision imaging systems development, integration and verification in military aircraft," Aerosp. Sci. Technol., vol. 31, no. 1, pp. 10-23, Dec. 2013. DOI: 10.1016/j.ast.2013.08.021.

[49]M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 32-64, Mar. 1995. DOI: 10.1518/001872095779049543

[50]D. Kahneman, *Attention and Effort*. Englewood Cliffs, NJ, USA: Prentice-Hall Inc., 1973, pp. 29-49.

[51]M. R. Endsley, "Predictive utility of an objective measure of situation awareness," in *Proceedings of the Human Factors Society Annual Meeting*, vol. 34, no. 1, Santa Monica, CA, USA, 1990, pp. 41-45. DOI: 10.1177/154193129003400110.

[52]K. J. Vicente, D. Craig and N. Moray, "Spectral analysis of sinus arrhythmia: a measure of mental effort," *Human Factors*, vol. 29, no. 2, pp. 171-182, Apr. 1987. DOI: 10.1177/001872088702900205.

[53]E. A. Byrne, and R. Parasuraman, "Psychophysiology and adaptive automation," *Biol. Psychol.,* vol. 42, no. 3, pp. 249-268, Feb. 1996. DOI: 10.1016/0301-0511(95)05161-9.

[54]M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proceedings of the National Aerospace and Electronics Conference (NAECON), New York, NY, USA,* 1988, pp. 789-795. DOI: 10.1109/NAECON.1988.195097.

[55]D. C. Logan, "Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry," *J. Exp. Bot.,* vol. 60, no. 3, pp. 712-714, Mar. 2009. DOI: 10.1093/jxb/erp043.

[56]M. D. Matthews, J. Eid, B. H. Johnsen, and O. C. Boe, "A comparison of expert ratings and self-assessments of situation awareness during a combat fatigue course," *Mil. Psychol*., vol. 23, no. 2, pp. 125-136, 2011. DOI: 10.1080/08995605.2011.550222.

[57]J. Mitzen, and R. L. Schweller, "Knowing the unknown unknowns: misplaced certainty and the onset of war," *Secur. Stud.,* vol. 20, no. 1, pp. 2-35, Mar. 2011. DOI: 10.1080/09636412.2011.549023.

[58]R. Pawson, G. Wong, and L. Owen, "Known knowns, known unknowns, unknown unknowns: the predicament of evidence-based policy," *Am. J. Eval.,* vol. 32, no. 4, pp. 518-546, Apr. 2011. DOI: 10.1177/1098214011403831.

[59]D. M. Gaba, S. K. Howard, and S. D. Small, "Situation awareness in anesthesiology," *Human Factors*, vol. 37, no. 1, pp. 20-31, Mar. 1995. DOI: 10.1518/001872095779049435.

[60]M. R. Endsley, "Situation awareness and human error: designing to support human performance," in *Proceedings of the High Consequence Systems Surety Conference*, Albuquerque, NM, USA, 1999, pp. 2-9.

[61]M. R. Houck, L. A. Whitaker, and R. R. Kendall, “A cognitive classification of pilot performance in air combat,” in *Proceedings of the IEEE 1992 National*, Dayton, OH, USA, 1992, pp. 503-509. DOI: 10.1109/NAECON.1992.220524.

[62]R. J. Lofaro, K. M. Smith, “Operational decision-making: integrating new concepts into the paradigm,” *in Proceedings of the Eleventh International Symposium on Aviation Psychology*, Columbus, OH, USA, 2000, pp. 1-6.

[63]U. Neisser, U. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. San Fransico, CA, USA: WH Freeman, 1976.

[64]K. L. Plant, and N. A. Stanton, N. “The process of processing: exploring the validity of Neisser's perceptual cycle model with accounts from critical decision-making in the cockpit,” *Ergonomics*, vol. 58, no. 6, pp. 909-923, Dec. 2014. DOI: 10.1080/00140139.2014.991765.

[65]T. J. Emerson, J. Reising, and H. Britten-Austin, “Workload and situation awareness in future aircraft,” SAE International, Tech. Rep. 871803, 1987. DOI: 10.4271/871803

[66]M. R. Endsley, and M. D. Rodgers, “Situation awareness information requirements analysis for en route air traffic control,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 38, no. 1, pp. 71-75. DOI: 10.1177/154193129403800113.

[67]A. R. Pritchett, R. J. Hansman, and E. N. Johnson, “Use of testable responses for performance-based measurement of situation awareness,” in *Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness*, Daytona Beach, FL, USA, 1995, pp. 75-81.

[68]G. Klein, “Analysis of situation awareness from critical incident reports,” in *Situation Awareness Analysis and Measurement,* M. R. Endsley, and D. J. Garland, Eds., Mahwah, NJ, USA: Lawrence Erlbaum, 2000, pp. 45-63.

[69]A. R. Pritchett, and R. J. Hansman, “Use of testable responses for performance-based measurement of situation awareness,” in *Situation Awareness Analysis and Measurement,* M. R. Endsley, and D. J. Garland, Eds., Mahwah, NJ, USA: Lawrence Erlbaum, 2000, pp. 170-190.

[70]A. J. Camm, et al., “Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology,” *Circulation*, vol. 93, no. 5, pp. 1043-1065, 1996. DOI: 10.1161/01.CIR.93.5.1043.

[71]H. ChuDuc, K. NguyenPhan, and D. NguenViet, “A review of heart rate variability and its applications,” *APCBEE Procedia*, vol. 7, pp. 80-85. DOI: 10.1016/j.apcbee.2013.08.016.

[72]E. R. Jones, R. T. Hennessy, and S. Deutsch, *Human factors Aspects of Simulation.* Washington DC, USA: The National Academies Press, 1985, pp. 69-76. DOI: 10.17226/19273.

[73]M. L. Cummings, L. F. Bertucelli, J. Macbeth, and A. Surana, “Task versus vehicle-based control paradigms in multiple unmanned vehicle supervision by a single operator,” *IEEE Trans. Human Mach. Syst.*, vol. 44, no. 3, pp. 353-361, Jun. 2014. DOI: 10.1109/THMS.2014.2304962.

[74]J. Zhang, Z. Yin, and R. Wang, R. “Recognition of mental workload levels under complex human–machine collaboration by using physiological features and adaptive support vector machines,” *IEEE Trans. Human Mach. Syst.*, vol. 45, no. 2, pp. 200-214, Apr. 2015. DOI: 10.1109/THMS.2014.2366914.

[75]S. G. Hart, and L. E. Staveland, “Development of NASA-TLX (Task Load Index): results of empirical and theoretical research,” in *Human Mental Workload,* P. A. Hancock, and N. Meshkati, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp.139-183.

[76]G. B. Reid, and T. E. Nygren, “The subjective workload assessment technique: a scaling procedure for measuring mental workload,” in *Human Mental Workload,* P. A. Hancock, and N. Meshkati, Eds. Amsterdam, The Netherlands: North-Holland, 1988, pp. 185-214.

[77]A. X. Miao, G. L. Zacharias, and S.-P. Kao, “A computational situation assessment model for nuclear power plant operations,” *IEEE Trans. Syst., Man, Cybern. A. Syst., Humans*, vol. 27, no. 6, pp. 728-742, Nov. 1997. DOI: 10.1109/3468.634636.

[78]R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Trans. Syst., Man, Cybern. A. Syst., Humans*, vol. 30, no. 3, pp. 286-297, May 2000. DOI: 10.1109/3468.844354.

[79]T. Haberkorn, I. Koglbauer, R. Braunstingl, and B. Prehofer, ”Requirements for future collision avoidance systems in visual flight: a human-centered approach,” *IEEE Trans. Human Mach. Syst.*, vol. 43, no. 6, pp. 583-594, Nov. 2013. DOI: 10.1109/THMS.2013.2284784.

[80]M. C. Dorneich, R. Dudley, E. Letsu-Dake, W. Rogers, S. D. Whitlow, M. C. Dillard, and E. Nelson, “Interaction of automation visibility and information quality in flight deck information automation,” *IEEE Trans. Human Mach. Syst.*, vol. 47, no. 6, pp. 915-926, Dec. 2017. DOI: 10.1109/THMS.2017.2717939.

[81]R. L. Keeney, and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge, England: Cambridge University Press, 1993.

[82]R. Clemen, and T. Reilly, *Making Hard Decisions,* 2nd ed., Belmont, CA, USA: Southwestern College Publications, 2004.

Heikki Mansikka is a retired Air Force fighter pilot (LtCol). He has a PhD from Coventry University, Aviation Human Factors and a MA from Kings' College London, Defence Studies. He is currently a Senior Manager, Pilot Performance in Etihad Airways, Abu Dhabi, UAE. His research concentrates on pilot performance.

Kai Virtanen is Professor of Operations Research at Department of Mathematics and Systems Analysis, Aalto University, Finland and at Department of Military Technology, National Defence University of Finland. His research interests include decision analysis, dynamic optimization, simulation-optimization and human performance in complex systems.

Don Harris is Professor of Human Factors at Coventry University. He is a Fellow of the Institute of Ergonomics and Human Factors and a Chartered Psychologist. Don is heavily involved in the design and development of the next generation of advanced pilot interfaces and the certification of civil flight decks.

.