Theoretical Issues in Ergonomics Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttie20

Measurement of team performance in air combat – have we been underperforming?

Heikki Mansikka, K. Virtanen, D. Harris & M. Jalava

To cite this article: Heikki Mansikka, K. Virtanen, D. Harris & M. Jalava (2021) Measurement of team performance in air combat – have we been underperforming?, Theoretical Issues in Ergonomics Science, 22:3, 338-359, DOI: 10.1080/1463922X.2020.1779382

To link to this article: https://doi.org/10.1080/1463922X.2020.1779382

THEORETICAL ISSUES IN ERGONOMICS SCIENCE
2021, VOL. 22, NO. 3, 338–359
https://doi.org/10.1080/1463922X.2020.1779382

Measurement of team performance in air combat – have we been underperforming?

Heikki Mansikka[a,b], K. Virtanen[c,d], D. Harris[e] and M. Jalava[f]

[a]Department of Mathematics and Systems Analysis, Aalto University, Helsinki, Finland; [b]Insta DefSec, Tampere, Finland; [c]Department of Mathematics and Systems Analysis, Aalto University, Helsinki, Finland; [d]Department of Military Technology, Finnish National Defence University, Helsinki, Finland; [e]Faculty of Engineering, Environment and Computing, Coventry University, Coventry, UK; [f]Satakunta Air Command, Pirkkala, Finland

ABSTRACT
In air combat, a traditional way of evaluating team's taskwork performance is to measure its performance output. However, it provides a narrow view about the team's performance and potentially misses the complexity of air combat and the team's taskwork when tactical operating procedures, teams' competencies and applicability of aircraft systems are evaluated in live-, virtual- and live-virtual-constructive-simulations. This paper introduces a model of an air combat system, which explains the dynamically interacting elements relevant to the measurement of team performance in air combat. Shared situation awareness, mental workload, task performance and normative performance are proposed as supplementary measures for the performance output. Recommendations for the selection of measurement techniques are provided. The multidimensional measurement approach presented in this paper can prevent potential misunderstanding of team performance. The principles of the system model and the associated measuring techniques can be applied to the analysis of any military system where the objective is to achieve a holistic measure of team performance.

ARTICLE HISTORY
Received 16 January 2019
Accepted 3 June 2020

KEYWORDS
Team performance; mental workload; normative performance; shared situation awareness; task performance

Relevance to human factors/ergonomics theory

This paper explains the elements relevant to the measurement of team performance in air combat and introduces shared situation awareness, mental workload, task performance and normative performance as supplementary measures for the team's performance output. This approach prevents potential misunderstanding of team performance in air combat and other military systems.

Introduction

When air combat simulations are used to evaluate and compare the utility of tactical operating procedures, the competence of teams or the applicability of aircraft systems, it is essential to have robust measures of team performance. While well-established measures

CONTACT Heikki Mansikka ✉ heikki.mansikka@aalto.fi Department of Mathematics and Systems Analysis, *Aalto University*, P. O. Box 11100, Aalto, FIN 00076, Finland.

of team and air combat performance exist, they tend to overlook some aspects of team performance and their interaction. Attempts to mechanically adopt the existing models of teamwork, such as the 'big five' (Salas, Sims, and Burke 2005), to air combat, have failed to address the degree of task or mission fulfilment and the quality of product delivery, and the way they interact with each other and with measures such as mental workload (MWL), performance output and situation awareness (SA) (see, e.g., Ohlander et al. 2019). To better understand performance in air combat, this paper considers air combat as a system, with necessary components to describe and measure team performance. Based on a model of this system, measures and measurement techniques for pilots' team performance are proposed to develop a holistic view of performance.

A team is 'a distinguishable set of two or more people who interact, dynamically, interdependently, and adaptively toward a common and valued goal/objective/mission, who have each been assigned specific roles or functions to perform, and who have a limited life-span of membership' (Salas et al. 1992). A team engages in taskwork and teamwork. Taskwork is activity related directly to performance in a task at hand, e.g., task goals and task performance requirements (Lim and Klein 2006; Marks, Zaccaro, and Mathieu 2000). Teamwork, on the other hand, deals with the team members' understanding and beliefs with regard to effective processes and understanding about the team itself (Mohammed, Klimoski, and Rentsch 2000; Rouse, Cannon-Bowers, and Salas 1992). This paper concentrates on the taskwork performance of a team.

A traditional way of evaluating a team's taskwork performance is to measure its performance output, which is the product of the taskwork. In air combat, the performance output is often measured as a ratio between the enemy aircraft shot down and friendly aircraft lost. However, a team's taskwork performance has more than one aspect and a sole performance output measure cannot capture all aspects. In air combat, like in any decision making activity, it is important to evaluate not just the product but also the process; it is possible to arrive in a right decision and still get the wrong outcome – or vice versa (Li and Harris 2006). To illustrate the different aspects of team performance, assume there are two teams, team A and team B. Both teams are tasked to follow different tactical operating procedures, typically referred to as tactics, techniques and procedures (TTPs). The utility of TTPs is evaluated based on the output performance of the teams. Let's assume both teams have similar performance outputs, e.g., no losses but also no enemy aircraft shot down. For several reasons, the team's performance output is not a powerful measure able to differentiate the utility of TTPs. First, while neither of the teams can complete the task of shooting down the enemy, assume that team A has more near kills, i.e., it is more effective in performing the process of shooting down the enemy. Similarly, while neither of the teams suffer any losses, assume that team A can also maintain a bigger safety margin, i.e., it is more effective in protecting itself, or performing the process of survival. Second, let's presume team A adheres to the directed TTP whereas team B does not. In such a situation, it is meaningless to even compare the utility of the directed TTPs. Third, for the team members to operate effectively, it is beneficial if they have a high level of awareness of the tactical situation (Carretta, Perry Jr, & Ree, 1996). Moreover, for the team members to operate effectively as a team, it is advantageous if the team members' awareness of the tactical situation is accurate and similar (Salas et al. 1995). Suppose the members of team A have more accurate and similar awareness of the tactical situation than the members of team B. Then, it can be argued that the team A is likely to be more effective if the complexity of the task is increased beyond the complexity of the original task. Fourth, team members possess limited cognitive resources to cope with the demands of the task. MWL measures the balance between the team members' cognitive resources and the task demands, and an unbalanced MWL will degrade team members' and team's performance (Mansikka et al. 2016). Assume the members of both teams have a balanced MWL, but the members of team A have a lower average MWL than the members of team B. While the members of both teams can satisfy the demands of the task, the members of team A have more spare cognitive capacity left to cope with the task demands should they exceed those experienced in the initial task. As this simplified example illustrates, a team's taskwork performance has more aspects than can be captured with just its performance output alone.

While there are more than 130 different models and frameworks of team performance (Salas et al. 2007), there are human factors/ergonomics concepts than are common between them. First, most human and team performance models and frameworks utilize an input-process-output (I-P-O) approach, or use different variants of it (Ilgen et al. 2005), to describe the fundamental phases of team performance (Hackman 1987; Kozlowski et al. 1999; McGrath 1984; Steiner 1972). Second, there is a wide consensus with regards to the role of affective (what teams feel), behavioural (what teams do) and cognitive mechanisms (what teams think) within the process phase (Grossman, Friedman, and Kalra 2017). The actual functions and components related to these mechanisms differ from approach to approach (see, e.g., Kozlowski and Ilgen 2006; Marks, Mathieu, and Zaccaro 2001; Salas, Sims, and Burke 2005). Despite the depth and breadth of the team performance literature, existing team performance measures typically consider just one or two aspects of performance, yielding a rather narrow view of the team's performance and missing the complexity of air combat and the team's taskwork.

For the purposes of building the air combat system model, this paper concentrates on the evaluation of decision-making and the essential affective and cognitive mechanisms related to it. More specifically, the system model focuses on shared SA and MWL as the central emergent states affecting inputs, team processes and the team performance outputs (Marks, Mathieu, and Zaccaro 2001). SA and MWL are commonly seen as the most essential concepts in team decision-making and human-system performance in complex systems (Parasuraman, Sheridan, and Wickens 2008; Tsang and Vidulich 2006; Vidulich and Tsang 2015; Wickens 2002). In fact, MWL and SA have become meta-measures for system evaluation (Selcon et al. 1996; Vidulich 2003). However, there is a need to evaluate SA, MWL and performance outputs separately, as there is a dissociation between them (Mansikka, Virtanen, & Harris, 2019).

When the objective is to evaluate and compare the utility of TTPs, the competence of the pilots or the applicability of aircraft systems, it is important to evaluate the teams' performance not just as a product (e.g., probability of kill, etc.), but also as a process. Air combat is a rapidly changing and uncertain environment; hence it is possible for the team members to apply sound tactical processes and still get a poor product – or vice versa. In other words, the team may exceed or fall short of a reasonable potential productivity baseline, i.e., can exhibit either process gain or loss (Kerr and Tindale 2004; Steiner 1972). The objective of the air combat system model proposed in this paper is to bridge the gap between traditional military performance measure (product) and the human-machine/human-human interaction (process) in a way that a holistic evaluation of team performance is possible. The air combat system model proposed, like any other model, is limited in the sense that it does not attempt to capture all aspects of the real-life phenomena it models. However, the system model introduced in this paper captures the most essential constructs affecting the processes underlying team's decision making; response selection and response execution; and also, the end product. By linking the 'what' with 'why', the model provides a tool to improve combat effectiveness (product) and gives an insight into human performance (process) in a context rarely accessible to researchers.

The system model is based on the existing principles of systems thinking and a number of well validated human factors concepts and measures. Although the model is slightly limited with regard to the performance output and decision-making aspects of the process, it brings together these principles and measures in a novel way in an air combat context. The model is a system model rather than a workflow model since in addition to the tasks and decisions involved in air combat, it also represents the evolution of the states of the combat situation over time, including the feedback phenomena related to these states.

The system model comprises two opposing air combat teams (i.e., the friendly team and enemy aircraft) which interact dynamically. Each performance output generated by either of the teams alters the whole system's state, and each change feeds back to the system causing further changes (Sterman 2001). The team's performance outputs are the outcomes of team processes (Kozlowski and Klein 2000), whereas the team's overall effectiveness is an assessment of the teams' performance outputs relative to some set of criteria (Hackman 1987). The performance outputs can result in irreversible state changes of the friendly team members and the enemy. In addition, the performance outputs can also result in transient state changes that dynamically change as the friendly team and enemy aircraft interact. Friendly team's processes, on the other hand, refer to activities that the team members engage in when they combine their resources to make decisions and to select responses which explain how or why the performance output is achieved (Kozlowski and Ilgen 2006; Smith, Borgvall, and Lif 2007). When undertaking these processes, the team members' information processing capacity may fall short of that required by the task demands resulting in excessive MWL (Svensson et al. 1997; Svensson, Angelborg-Thanderz, and Sjöberg 1993). An unbalanced MWL has a negative effect on shared SA (Endsley 1995; Svensson and Wilson 2002), which in turn is an essential precursor to decision making, response selection and eventually the effectiveness of team's performance outputs (Endsley 1993; Houck, Whitaker, and Kendall 1993; Waag and Houck 1994). The air combat model described in this paper is constructed such that it captures both the irreversible and transient state changes reflecting the friendly team's performance output. In addition, it reflects team's performance processes by incorporating both MWL and shared SA in the model (Mansikka, Harris, and Virtanen 2019b). However, the linkage between MWL, SA and performance output is not direct (Mansikka, Virtanen, & Harris, 2019). Despite a satisfactory SA and a balanced MWL, the pilots may arrive at (seemingly) poor decisions, which do not coincide with those dictated by the tactical operating procedure most suitable for the task at hand. In addition, the pilots can arrive in a rational decision, but fail in their response execution. As a result, to capture the quality of friendly team's response execution, the team's TTP adherence is also incorporated in the model.

The model is primarily concerned with the friendly team's performance output and the processes underlying the quality of team's response execution and the main precursors of team's decision-making and response selection. To support a holistic evaluation of team's performance, measures of each of these are included in the model. In addition, recommendations for the selection of team performance measurement techniques are provided when the TTPs, teams' competencies and applicability of aircraft systems are evaluated in live- (L), virtual- (V) and live-virtual-constructive- (LVC) simulations. In a L-simulation, real people operate real systems. In a V-simulation, real people operate simulated systems or simulated people operate real systems, and in a C-simulation, simulated people operate simulated systems (Hodson and Hill 2014). In a LVC-simulation, the different simulation classes are mixed. Each simulation class has its strengths and weaknesses. A C-simulation allows a precise control of the simulation entities, but the lack of real people makes it challenging to evaluate human-machine interaction. With real people making decisions, a V-simulation is a practical way of evaluating, e.g., MWL and SA – especially if optimal TTPs are first determined in a C-simulation. A L-simulation is an expensive and resource heavy but allows taking real-life task complexity and stressors into account. As such, the L-simulation should be used only after the potential safety issues have been mitigated in C- and V-simulation. While LVC-simulation is powerful method to evaluate large-scale systems and systems of systems, it is limited by known interoperability and architectural issues (Zhihua et al. 2013).

System model

Structure

The BVR air combat system is comprised of assets such as aircraft, surface-to-air missile systems, radar sites and command and control centres, together with the personnel who operate these assets. In BVR air combat, the friendly team uses its airborne detection equipment to search for the enemy aircraft and employs remote air-to-air missiles to attack them while staying beyond the visual range from the enemy (Paddon 1977). For the sake of brevity, the system model introduced in this section discusses the beyond visual range (BVR) air combat between a single friendly team (a flight of four aircraft), enemy pilots and their assets.

The system model identifies three types of actors. An actor represents a person or group of persons who undertakes an activity described in the system model. First, the combat agents include the friendly team members and enemy pilots. Second, the ground truth holder is an actor that has access to the objective reality, i.e., the real states of air combat agents and their assets. A simulation supervisor is a typical ground truth holder. The real states refer to the states of the air combat agents as they really are - as opposed to the states they are thought to be in, i.e., their observed states. A real state describes whether a combat agent is dead or alive and the current phase of taskwork. Third, a measurer is an external observer who measures the friendly team's performance. A measurer is a person with necessary air combat subject matter expertise and who is trained to use the measurement techniques discussed later in this paper.

The US Joint Doctrine describes taskwork as two parallel processes, typically referred to as chains, with six phases in each (Joint Chief of Staff 2013). The phases in the other chain describe the progression of friendly team members' taskwork, whereas the phases in the other chain describe how well the friendly team members can deny the enemy's taskwork from progressing. The selection and execution of air combat agents' taskwork related responses change their real states. The friendly team's objective is expressed as the combat agents' desired real states, and the friendly team's task is to minimize the difference between their objective and the real states. However, the air combat agents don't have access to the real states during flight and must rely on the observed states when determining the progression of air combat. Therefore, the observed states are the estimates of the real states. The friendly team's objective is typically given in an air tasking order (ATO). In addition to objectives, the ATO lists the friendly teams and assets available for the mission and allocates the tasks to them.

In the following sub-sections, the description of the air combat system model is divided between the actors from the three groups. The superscripted numbers in these sub-sections refer to the elements of the air combat system illustrated in Figure 1. Both the sub-sections and Figure 1 are constructed from the friendly team's perspective and the measures discussed deal with the performance of the friendly team. Solid lines in Figure 1 refer to the activities or states of the friendly team members. The dashed lines in Figure 1 refer to team activities, states or performance measures.

Ground truth holder

Changes in the air combat agent's states update the real states of the friendly team[1] and the enemies.[2] The updated states result in a real state of air combat formed.[3] The ground truth holder will observe the real state of air combat[4] and form a real state of air combat observed.[5] The ground truth holder's observations are unique in a sense that they do not have any error.

Air combat agents

As the friendly team has many concurrent tasks, it must allocate the cognitive resources[6] possessed by the team members. The team members start to consume their cognitive resources for the taskwork as soon as they begin to observe the real states of the friendly team[7] and of the enemy.[8]

When the friendly team members interact with the air combat system, they cannot truly know its real state, i.e., the objective reality, which is external to their senses. Instead, using their pre-existing knowledge and observations about the elements of the system, the team members form cognitive representations about the system and its elements, i.e., they develop individual subjective realities. The knowledge that forms the foundation of the team members' subjective reality is held in their long-term memories (LTM) and is relatively stable. LTM held knowledge is organized in the form of semantic networks, with each section of the network containing knowledge related to a specific element of the system (Wickens et al. 2004). The knowledge content and structure about the system and its elements are referred to as schemas (Rumelhart 1984; Rumelhart and Ortney 1977), whereas scripts depict the appropriate sequence of activities and behaviours related to a specific task (Wilson and Rutherford 1989; Schank and Abelson 1977, 1988). Both scripts and schemas are context specific in a sense that each is associated with a particular system element within the task environment (Wickens et al. 2004; Langan-Fox, Code, and Langfield-Smith 2000). The scripts and schemas which form the team members' subjective reality of a dynamic system are collectively called mental models (MMs) and contain pilots' knowledge regarding the team and the enemies, together with the knowledge about the sequence of activities regarding the team's task (Gilbert 2011; Wilson and Rutherford 1989). The

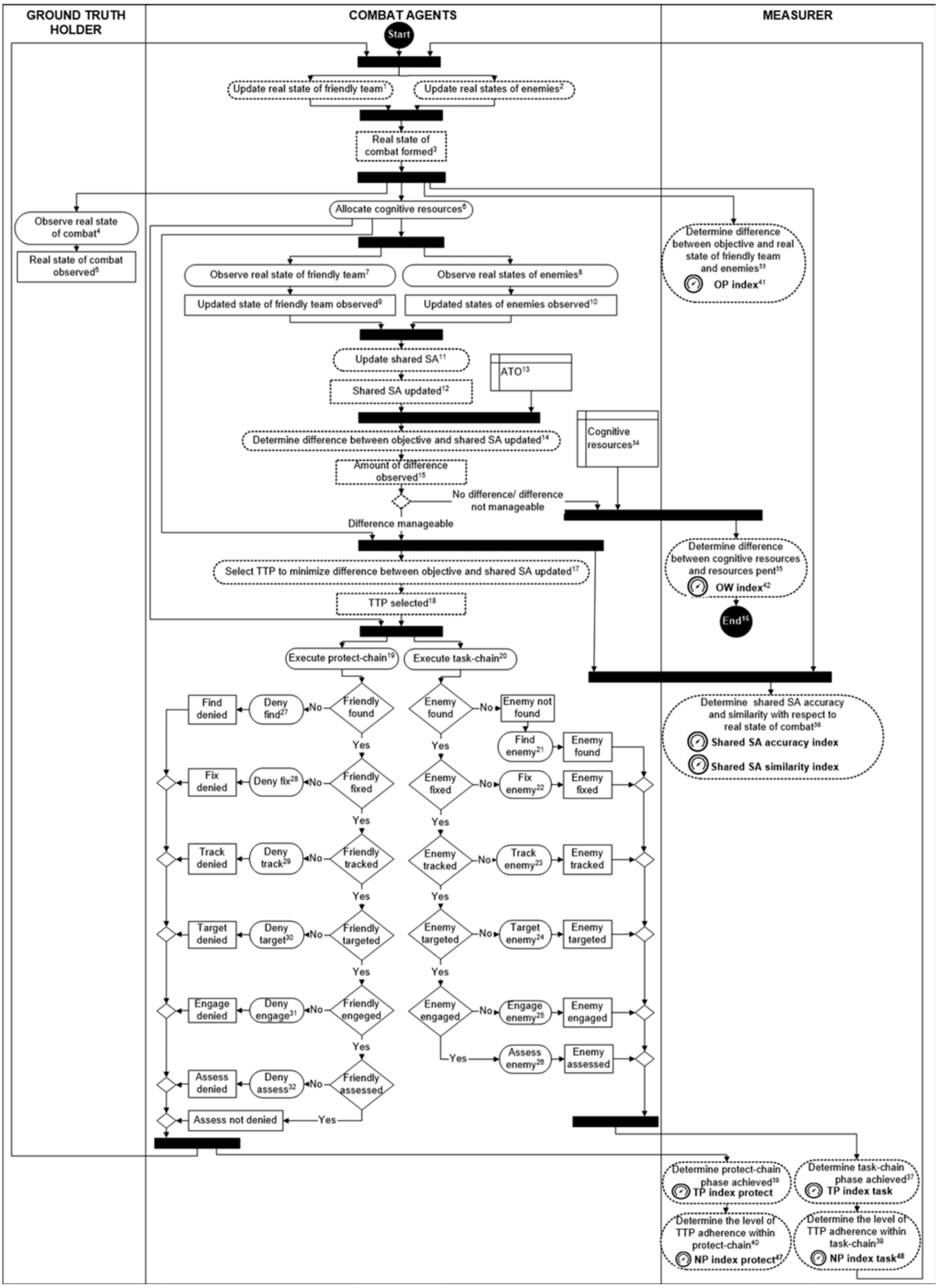

Figure 1. Air combat system model and performance measures from the perspective of the friendly team. Solid lines refer to the activities or states of the friendly team members and the dashed lines refer to team activities, states or performance measures.

friendly team members utilize their initial mental models (MMs) about the enemy as well as about their own team and its task when directing their attention to observation, and when updating their own mental models (Neisser 1976). MMs are internal representations, which include state estimates of the real states of the team and of the enemy. Based on team members' observations, each of them forms updated state estimates of the friendly team[9] and the enemy.[10]

The internally held mental models exist even without interaction with the system. When an interaction takes place, the mental models assist exploration of their environment by guiding attention allocation and enable interpretation of the perceived information in a way that allows an individual to understand its relevance in relation to task and task goals (Baxter, Besnard, and Riley 2007; Endsley 1995; Neisser 1976). Finally, using the knowledge obtained through interpreted perceptions and experience in the form of mental models, a person can run forward-looking mental simulations to forecast future events and system states (Endsley 1995; Rouse and Morris 1986; Salmon et al. 2008, 2009). These dynamically updated mental models can be understood as activated knowledge about a situation in which one is currently involved, also known as situation awareness (SA) or situation models (Endsley 1995; Endsley 2000; Langan-Fox, Anglim, and Wilson 2004; Wickens et al. 2004).

While there is no universally agreed definition of SA (Salmon et al. 2006), it is typically represented as an operator's dynamic understanding of 'what is going on' (Endsley 1995). Equally, the theoretical approaches underpinning SA vary greatly, see, e.g., Endsley (1995), Smith and Hancock (1995), Bedny and Meister (1999), Taylor (1990), and Sarter and Woods (1991). Of the different approaches, Endsley's three-level model, which has received most attention, defines SA as 'the perception of environmental elements and events with respect to time or space (SA level 1), the comprehension of their meaning (SA level 2), and the projection of their future status (SA level 3)'. In Endsley's model, each SA level is built upon the level below such that a low SA at a lower level may contribute to a low SA at a higher level (Endsley & Robertson, 2000; Endsley and Garland 2000; Endsley and Jones 2001). For example, a failure in perception (SA level 1 error) can lead to a lack of or incomplete mental model (SA level 2 or SA level 3 error) (Jones and Endsley 1996).

While SA was originally introduced as an individual level construct, it has increasingly been used to describe how teams (Endsley and Jones 2001; Endsley & Robertson, 2000; Saner et al. 2009) and entire socio-technical systems (Stanton et al. 2010; 2009; 2006) become and remain coupled to the their operational environment (Moray 2004). Shared SA describes the degree to which there is an overlap of the individual SA between team members on mutual operational requirements.

The quality of team members' SA of shared elements (as a state of knowledge) may serve as one index of a team's effectiveness in a human-machine system (Endsley 1995). As such, shared SA essentially describes the team members' shared and organized understanding of relevant knowledge about the system they are interacting with (Cannon-Bowers, Salas, and Converse 1993; Klimoski and Mohammed 1994). Shared SA has two aspects: similarity and accuracy. Similarity refers to the extent to which the team members' individual MMs are in an agreement with one another (Mathieu et al. 2005), whereas accuracy of shared SA is the level of agreement between the team members' MMs, and the real state of the team and the enemies. The accuracy and similarity of shared SA are essential contributors to task performance as they help the team members to coordinate their actions (Converse, Cannon-Bowers, and Salas 1991; Langan-Fox, Anglim, and Wilson 2004), predict each other's behaviours and needs (Jonker et al. 2010) and to interpret the objective reality in a similar manner (Johnson et al. 2007).

An initial shared SA includes the team's estimates of the real states of the team and the enemies, which are formed from team members' individual state estimates. A team utilizes the mechanisms and processes of teamwork, mainly communication and coordination, to update shared SA.[11] As a result, their initial shared SA is converted into shared SA updated.[12] The team uses its shared SA updated and the objective given in the ATO[13] to determine the difference between the objective and shared SA updated.[14] The subjectively comprehended difference between the objective and shared SA updated dictates the degree of difference observed.[15] If there is no difference, the team's task has been completed and its taskwork reaches an end[16] Also, the team will abort its mission if the observed difference between objective and shared SA updated is not manageable. If, however, the observed difference is manageable, the team will begin to intercept the enemies of interest. Whether shared SA updated is accurate and similar or not, the team uses it to select TTP to minimize the difference between the objective and the shared SA updated.[17] The TTP selection is ultimately the decision of the team leader. However, here the TTP selection is considered as a team activity as all team members can recommend TTP to be selected and, time permitting, the team can discuss the TTP options.

The implementation of TTP selected[18] requires management of two parallel chains; the team members must execute both their protect[19]- and task-chains.[20] There is a separate task-chain for every enemy aircraft to be intercepted. The team members can manage many task-chains simultaneously and they can manage the task-chains either independently or jointly. In comparison, every team member has its own protect-chain. The task-chain is comprised of a sequence with six phases. In BVR air combat, the enemy is found and identified in the Find-phase.[21] In the Fix-phase,[22] enemy's exact location is obtained. In BVR air combat, this phase is often automated. In the Track-phase,[23] the enemy is monitored until a decision is made to either engage or to discontinue the commit. In the Target-phase,[24] a weapon or a sensor is assigned against the enemy. In the Engage-phase,[25] the assigned weapon or sensor is employed against the enemy. Finally, in the Assess-phase,[26] the state of the enemy is assessed, and any required follow-on actions are determined.

Force-on-force air combat is a zero-sum game. Both the team members and the enemies have similar task- and protect-chains, where a gain of the team member results in an equal loss of the enemy and vice versa. The protect-chains are essentially about denying the opponent's task-chains from progressing. Therefore, the protect-chains mirror the opponent's task-chains; while both sides attempt to complete their task-chains, they also try to keep their protect-chains intact, i.e., they attempt to deny the opponent from progressing their task-chains. Due to this equilibrium of the chains, the protect-chain has similar phases[27-32] and sequence to the task-chain. A phase change in a task-chain will result in a phase change in a respective protect-chain – and vice versa. Each time the phase changes are completed, the real states of the team member and the enemy are updated to reflect the new chain phases. Therefore, the air combat agent's state tells whether the agent is dead or alive, and at what phase its task and live chains are. For example, assume a BVR air combat between a friendly team member and an enemy pilot where both pilots are alive, the friendly team member has engaged the enemy pilot and the friendly team member has not been found by the enemy pilot. In such situation, the friendly team member would be in a state "alive, enemy engaged, find denied." At the same time, the enemy pilot would be in a state "alive, enemy not found, assess denied." It should be remembered, however, that the updated mental model of the pilot and the ground truth may be in disagreement; a team member may believe it has not been detected while it has, or it may think that a correct enemy has been targeted when it has not.

Measurer

Output performance (OP) is probably the most traditional measure of task performance in air combat. OP measures the difference between the objective and the real state of the team and the enemies relative to some set of criteria.[33]

The team members' cognitive resources are taxed throughout its taskwork. A mismatch between the friendly team members' cognitive resources[34] available and the resources spent can generate an unbalanced MWL, which can eventually degrade performance. Therefore, to gain a better understanding of the factors affecting taskwork performance, it is useful to determine the difference between the team members' cognitive resources and the resources spent[35] during the perceptual encoding, central processing, response selection and response execution stages.

Although shared SA is continuously updated with the updated MMs and teamwork, shared SA accuracy and similarity may be low. As the team selects TTP with whatever shared SA it has, it is insightful to determine shared SA accuracy and similarity with respect to real state of air combat,[36] as such an evaluation reveals how well-informed the team was when deciding to intercept the enemy and when selecting TTP.

It is possible that teams with different pilots, aircraft systems or operating principles will achieve an equal OP when engaged in a similar air combat situation. This does not mean, however, that their taskwork performances would be similar in terms of task performance (TP), i.e., the progression of the team members' task-chains and the guarding of their protect-chains. If TPs are dissimilar while OPs are similar, the team whose members' protect-chains are more intact, has maintained a larger margin of survival. Similarly, a team whose members' task-chains have progressed further has been closer to achieving the desired effect on the enemy. Therefore, in addition to the OP measure it is informative to measure task- and protect-chain TPs, i.e., to determine task-chain phase achieved[37] and to determine protect-chain phase achieved[38] by evaluating how far the task-chains progressed and how complete the protect-chains were maintained.

Each TTP consists of a set of quantitative and qualitative rules, which allow the team to coordinate the taskwork among its members (Mansikka, Harris, and Virtanen 2019a). A quantitative rule has a variable and its value. For example, 'Airspeed at missile launch must be Mach 1.0′ is a quantitative rule, where Mach 1.0 is its value. A qualitative rule consists of a verbal description of activity. For example, 'Team members must communicate their tactical status' is a qualitative rule. A TTP rule may be linked to a specific phase of protect- or task-chain, or it may be associated to many phases. With respect to a team member engaged in air combat, normative performance (NP) represents whether the team member adheres to the rules of the selected TTP. To evaluate whether it is even sensible to compare the team members' effectiveness, TTPs or aircraft equipment, it is necessary to determine the level of TTP adherence within task-chain[39] and to determine the level of TTP adherence within protect-chain.[40] This approach has been used to assess pilot's performance in a BVR air combat (Mansikka, Harris, and Virtanen 2019b).

Recommendations for measurement techniques

The measurement techniques described are meant to be used after the L- or V- simulations. In the following sub-sections, a term 'score' refers to an interim result, whereas a term 'index' refers to the final result.

OP measuring techniques

The objective of the team dictates the most appropriate OP measure(s) to evaluate task accomplishment. The team may be expected to deliver either kinetic or non-kinetic effects - or both. If the objective is to produce non-kinetic effects, it may be enough if the team just shows its presence in an area of responsibility. Typical measures of kinetic OP include the kill-loss ratio and the number of missiles or time required to shoot down the enemy. The selection of OP measures is typically straightforward as the attributes of OP are normally provided by the military staff. Depending on the measurement objective, OP can be determined when a state change occurs or at certain time intervals. OP score is measured for each team member. An average of OP scores is used as the team's OP index.[41] OP can be measured during a standard debrief. The objective of the debrief is to determine if the desired mission objectives were achieved, identify lessons learned, and define aspects of training needing improvement (Royal Korean Air Force (RoKAF) 2005). In the debrief, the previously flown mission is reconstructed and reviewed by all mission participants, including the friendly team members and flight instructors.

MWL measuring techniques

MWL can be measured using performance, subjective, and physiological measuring techniques (O'Donnell, Eggemeier, and Thomas 1986). Physiological measures are often poorly suited for L-, V- or LVC-simulations as they are non-diagnostic, require disruptive instrumentation, and are sensitive to second- and third order physiological effects and environmental factors (Lutfi and Sukkar 2012). Performance measures, on the other hand, assume that the team member's primary or second task performance is related to workload (Paas et al. 2003). The secondary task measures cannot be safely utilized in L-simulations if the expected mental workload is high (Casali and Wierwille 1984) and some argue that the primary task measures should not be used as workload measures at all (Hart and McPherson 1976; O'Donnell, Eggemeier, and Thomas 1986).

Measuring techniques utilize the team members' subjectively experienced workload, i.e., how a person feels when doing a task (Johanssen et al. 1979). Among the numerous subjective measuring techniques available (see, e.g., Reid and Nygren 1988; Roscoe and Ellis 1990; Wierwille and Casali 1983), NASA Task Load Index (NASA-TLX) (Hart and Staveland 1988) has gained the widest acceptance, mainly as it is non-intrusive and easy to implement (Mansikka, Virtanen, & Harris, 2019). NASA-TLX identifies six MWL dimensions, which the pilots score separately from 0 (low MWL) to 100 (high MWL). The dimensions are:

Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Own Performance (OP), Effort (EF), Frustration (FR). A team's average MWL score for each dimension is calculated by averaging the team member's dimension scores. An overall workload (OW) score for each team member is calculated by averaging the dimensions' scores, whereas an OW index[42] is the mean of the team members' OW scores. While NASA-TLX has some known issues related to potential pilot bias and limitations of memory recall, it is the most suitable technique to measure MWL in air combat – at least when used in a non-punitive setting. To minimize a retrospective bias related to NASA-TLX (Young et al. 2015), it is recommended that the MWL measurement is conducted immediately after the simulation. It should be noted that NASA-TLX provides an ex post-facto measure of the average MWL. Therefore, MWL indexes or scores cannot be presented as a function of time.

Shared SA similarity and accuracy measuring techniques

Shared SA similarity and accuracy can be measured in several ways, e.g., observer rating, freeze probe, and self- and peer appraisal techniques (Sulistyawati, Wickens, and Chui 2009). While appraisal techniques are low-cost and easy to administer, pilots may be unable to accurately report what they and their peers are and are not aware of (Clark and Elen 2006; Feldon and Clark 2006). Freeze-probe techniques, on the other hand, are far too intrusive to be used in L-, V- or LVC-simulations (Stanton et al. 2013). Finally, as the observer rating techniques rely on observable behaviours as indicators of shared SA, there are doubts to what extent an external observer is able to assess a pilots' subjective reality which is not necessarily manifested in overt behaviour.

An indirect option to evaluate shared SA similarity and accuracy is to elicit the similarity and accuracy of team members' selected MMs. While most MM elicitation techniques are too time-consuming for operational use (Langan-Fox, Code, and Langfield-Smith 2000), retrospective verbal probing (RVP) (Beatty and Willis 2007; Schaeffer and Presser 2003; Willis, Royston, and Bercini 1991), a sub-category of cognitive interview techniques (Fisher and Geiselman 1992; Fisher, Geiselman, and Amador 1989; Geiselman et al. 1986; Willis, Royston, and Bercini 1991) is a promising technique to evaluate shared SA similarity and accuracy in a natural air combat training environment. In the RVP technique, the team members are engaged in a facilitated, semi-structured conversation during the debrief to elicit the accuracy and similarity of their MMs. Shared SA similarity and accuracy is then determined based on the similarity and accuracy of team members' MMs.

The knowledge relevant to each MM is determined before the debrief. The relevant knowledge is broken down to smaller units, or concepts. A single probe is developed to tap each concept. Due to the complexity of MMs and associated knowledge, there can be several concepts associated to a single MM. The probes are written as questions. For example, the team members have a MM regarding an enemy, with enemy's offensive capability being one concept related to that MM. The associated probe would then be: "Did you know the offensive capability of the enemy?" With properly selected concepts and probes, the RVP technique allows capturing team members' understanding of complex and abstract MMs, such as awareness about the team members' awareness.

The elicitation procedure is similar to a standard debrief, where the pilots routinely recall their understanding of the situation and compare it with the real states available in the debrief (Kelly et al. 1979; Lilja, Brynielsson, and Lindquist 2016; Waag and Houck 1994). As the MMs are elicited, a flight instructor or team leader acts as a facilitator and pauses the mission replay at certain time intervals or after an event of interest has occurred. Then, the facilitator identifies which concepts were relevant to the team since the previous pause. The facilitator introduces each concept one by one and initiates RVP to elicit them. As a typical mission reconstruction includes radio communications, flight trajectories and cockpit recordings of all mission participants, all pilots and instructors attending the debrief have access to the objective reality.

Protect- and task-chain TP measuring techniques

Protect- and task-chain TP are useful measures when the objective is to differentiate teams' taskwork performance when there are no significant differences in their OP indexes. Before TP can be measured, unambiguous indicators of the completion of each phase of the task- and protect-chains must be identified. For example, a missile launch against an aircraft is an unambiguous indicator that the target aircraft has been engaged by the launch aircraft. Evaluating TP accurately can be rather time consuming as the determination of each task- and protect-chain phase completion requires detailed review of the friendly team and enemy states. Therefore, if TP is attained during the debrief, it is advisable to determine a snapshot of TP at certain time intervals instead of evaluating every phase. If, on the other hand, TP is determined after the debrief, all TP indicators must be selected such that TP can be obtained without team members' assistance simply by just reviewing the mission reconstruction and team members' cockpit recordings.

NP measuring techniques

To measure NP, relevant TTP rules must be first identified for the TTP of interest. NP measurement can be too time consuming to be conducted during the debrief as it requires detailed analysis of the pilot's cockpit recordings. Therefore, both qualitative and quantitative rules should be selected such that a measurer (typically a flight instructor) can determine the team members' TTP adherence after the debrief and without their assistance.

A NP score is determined for each TTP rule and team member. The NP score for a single TTP rule is obtained by aggregating the team members' TTP adherences of that rule. Each team member's individual NP score is formed by aggregating the TTP adherence of all TTP rules that apply to that team member. The most straightforward way to determine the NP scores is to give each correctly executed TTP rule of interest a value of '1′ and each wrongly executed or omitted TTP rule a value of '0′. However, sometimes it may be reasonable to evaluate NP with respect to team's ability to execute the TTP. If that approach is preferred, each correctly executed TTP rule would be given a value of '2′ and an omitted or wrongly executed TTP rule without a significant impact on TTP execution would be given a value of '1′. Finally, an omitted or wrongly executed TTP rule with a significant impact on TTP execution would be given a value of '0′.

NP indexes are obtained by aggregating the individual NP scores. The NP scores and the NP indexes[47,48] can be determined separately for each task- and protect-chain phase, for complete chains or for the TTP as a whole. The high pace of air combat and the number of task-chains may not encourage determining the NP index for every task- and protect-chain phase after each phase change.

Discussion

In an air combat, a flight faces complex multi-objective decision problems with potentially disastrous impacts on team performance. Previous research has traditionally utilised individual team performance outputs (e.g., accuracy, quality), and has less frequently provided a blended or composite measure of team outputs (Mathieu et al., 2008; Sundstrom, De Meuse, and Futrell 1990). The degree of task or mission fulfilment, however, has received no attention in air combat domain, and very little interest in other domains (Van Der Vegt and Bunderson 2005). This study contributes to existing literature by explicitly demonstrating how, and why, a task fulfilment measure (i.e., TP) increases the diagnosticity of the measurement of team performance output. To understand the performance of a team, one must model it in its context (Goodman, Ravlin, and Schminke 1987). When such modelling attempts have been made to understand air combat, the work has merely translated some existing models of teamwork (cf. Salas, Sims, and Burke 2005) into aviation language (see, e.g., Ohlander et al. 2019). In addition, while the team performance literature acknowledges product quality as one possible performance product of a team, this study is more specific by introducing the quality of product *delivery* (i.e., NP) as an additional and valuable measure of team performance. Without such a measure, using team performance measures to compare the effectiveness of processes or procedures may be arbitrary. This notion should expand the general team performance interest of "what teams do" from process phase to output phase. Overall, this study contributes to existing literature by not only modelling air combat team's performance in its context, but also by considering performance aspects relevant in that specific context.

This paper described a holistic way of measuring team performance in air combat when the utility of TTPs, the competence of teams or the applicability of aircraft systems is evaluated and compared in L- and V- simulations. In addition to traditional OP measures, MWL, shared SA similarity and accuracy, NP and TP were introduced as additional team performance measures. That said, there are no global thresholds or redlines for OP, TP, NP, MWL or shared SA, and they can be considered either as criteria or constraints. Therefore, it may be practical to consider the evaluation and comparison task as a multicriteria decision analysis problem (Clemen and Reilly 2004; Keeney and Raiffa 1993) where the criteria and the constraints are defined by a decision maker. The model introduced in this paper is the first time air combat is described as a system model containing several team performance measures allowing multi-dimensional analysis of team performance in air combat setting. As such, the model offers a framework to analyze the flight's performance in different air combat scenarios by combining well-established human factors principles and methods in a novel way with the principles of systems thinking and the fundamentals of air combat.

All models are a simplification of reality and are highly dependent on the quality of data obtained. The utility of the model, though, can be determined by its usefulness in adding further explanatory detail to what initially appears to be both good and poor performance. In an uncertain environment, good decisions will not always result in good outcomes. However, by better understanding the system-wide relationship between performance and the cognitive factors in team members, potential areas of improvement may be identified. Small gains in SA or MWL at the right time may be related to significant performance improvements. Conversely, what may initially be expected to be large gains in output related to significant gains in SA or MWL may prove not to have any effect as the rate-limiting step lies elsewhere in the system. Only by understanding the whole system can performance ultimately be understood. In an air combat situation, potential performance gains, and cognitive benefits (through increased SA and or decreased MWL) may be gained by the introduction of new equipment (interfaces or sensors) or TTPs (Mansikka, Harris, and Virtanen 2019b). A group of subject matter experts were closely involved in developing the model in order to verify that the model captures the relevant aspects of air combat such that the finalized model remains relevant for its practical use.

Overall, the model of air combat described holds a great deal of promise for improving the understanding of team decision making and performance in air combat. In more general terms, the principles of the system model and the associated measuring techniques should serve as a sound point of departure for the analysis of any military system where the objective is to achieve a holistic measure of team performance. Although the model is predicated upon team/task working in a flight of fighter aircraft, the basic principles underlying the model hold good for any team undertaking a coordinated activity in a dynamic, uncertain environment. For example, business decisions taken by Board members are also undertaken in an uncertain environment where the ground truth is uncertain and the quality of the teamwork cannot be evaluated by the outcome alone. The only difference in this instance is in the nature of the underlying information and the speed of the evolving situation. The air combat situation is, however, unique on the demands that it places on the participants, which have serves to drive the construction of this model of teamworking.

While this paper concentrated on the measurement of the team's taskwork output and teamwork results in a flight of fighter aircraft, further research of team processes (Mansikka, Harris, and Virtanen 2019; Marks, Mathieu, and Zaccaro 2001) is required to understand how the team arrives on these results. The next steps will be to further refine the model and the measurement base to support quantitative modelling of the relationships between actors and assets (for example, by using a structural equation modelling-based approach). This will also allow new parameters to be introduced into the model in a systematic manner and permit the understanding of their scientific contribution. The contribution of any variable in such a situation can only be evaluated by trading off its utility (in terms of model comprehensiveness and prediction) against its added level of complexity: start simple and add complexity, as required.

As a first step, the air combat system model will be validated in an empirical setting to see if it is possible to collect the required data in an operational environment. While it may be intuitively tempting to test the model in a C-simulation, such test design would be challenging as MWL, SA and other human-machine interaction aspects would have to be modelled – as opposed to measuring the cognitive responses and reactions of real pilots. Due to the nature and type of team modelling required (i.e., combat aircraft and/or virtual simulators and operational fighter pilots), using V- or L-simulations to validate the model is not a simple task. There is a risk that the pace of operational flight training will not enable the data collection in a way the model requires.

Although the air combat system model is based on well-established cognitive constructs, systems thinking principles and air combat standards, the model is illustrated as a rather simplified graphical representation. This is a value in itself, as it allows the model to be understood by air combat professionals - possibly with little human factors or systems thinking background. In fact, operational fighter pilots have already evaluated the principles of the proposed model, and the initial responses suggest that the model is understandable, and the data collection should be possible in an operational setting. Moreover, some pilots involved in the development and evaluation of the model have expressed that it has helped them to obtain new insights into team performance in air combat already at this pre-implementation phase.

The authors are currently preparing an empirical study where the air combat system model is validated using V- and L-simulations and real fighter pilots. The empirical study will be built on the success of previous studies where NP, MWL, SA and OP have been measured in air combat simulations (Mansikka, Harris, and Virtanen 2019a; Mansikka, Harris, and Virtanen 2019b; Mansikka, Virtanen & Harris, 2019; Mansikka, Virtanen & Harris, 2019). In the planned study, MWL will be evaluated with NASA-TLX and shared SA will be evaluated with RVP. TP will be determined by evaluating how intact the flight members can maintain their protect chains and how far their task chains progress against the enemy aircraft. NP will be determined based on the pilot's adherence of operational TTPs. Finally, OP will be determined using actual measures of task performance provided by the military subject matter experts. Once the model has been validated in an operational environment, the objective is to implement it in day-to-day training, research and development, and system evaluations, where it can help to draw a holistic picture of flight's performance in air combat.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Heikki Mansikka is a retired Air Force fighter pilot (LtCol). He has a PhD from Coventry University, Aviation Human Factors and a MA from King's College London, Defence Studies. He is currently a Chief Air Combat Scientist at Insta DefSec. His research concentrates on fighter pilots' performance and air combat.

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.

Matti Jalava has held various research and development posts within defence industry and Finnish Defence Forces since 1997. He holds a MSc from Helsinki University of Technology. He is currently serving in Satakunta Air Command, Pirkkala, Finland.

ORCID

Heikki Mansikka http://orcid.org/0000-0002-7261-6408

References

Baxter, G., D. Besnard, and D. Riley. 2007. “Cognitive Mismatches in the Cockpit: Will They Ever Be a Thing of the Past?” *Applied Ergonomics* 38 (4): 417–423. doi:10.1016/j.apergo.2007.01.005.

Beatty, P., and G. Willis. 2007. “Research Synthesis: The Practice of Cognitive Interviewing.” *Public Opinion Quarterly* 71 (2): 287–311. doi:10.1093/poq/nfm006.

Bedny, G., and D. Meister. 1999. “Theory of Activity and Situation Awareness.” *International Journal of Cognitive Ergonomics* 3 (1): 63–72. doi:10.1207/s15327566ijce0301_5.

Cannon-Bowers, J., Salas, E., & Converse, S. 1993. Shared mental models in expert team decision making. In *Individual and Group Decision Making*, edited by J. Castellan, 221–246. Hillsdale, NJ: Lawrence Erlbaum.

Carretta, T., D. Perry, Jr, and M. Ree. 1996. “Prediction of Situational Awareness in F-15 Pilots.” *The International Journal of Aviation Psychology* 6 (1): 21–41. doi:10.1207/s15327108ijap0601_2.

Casali, J., and W. Wierwille. 1984. “On the Measurement of Pilot Perceptual Workload: A Comparison of Assessment Techniques Addressing Sensitivity and Intrusion Issues.” *Ergonomics* 27 (10): 1033–1050. doi:10.1080/00140138408963584.

Clark, R., and J. Elen. 2006. “When Less is More: Research and Theory Insights about Instruction for Complex Learning.” *Handling Complexity in Learning Environments: Research and Theory*, edited by R. Clark and J. Elen, 283–297. The Netherlands: Elsevier.

Clemen, R., and T. Reilly. 2004. *Making Hard Decisions: An Introduction to Decision Analysis*. 2nd ed. Belmont, CA: Southwestern College Publications.

Converse, S., J. Cannon-Bowers, and E. Salas. 1991. “Team Member Shared Mental Models: A Theory and Some Methodological Issues.” *Proceedings of the Human Factors Society Annual Meeting* 35 (19): 1417–1421. doi:10.1177/154193129103501917.

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. doi:10.1207/s15327108ijap0302_5.

Endsley, M. 1995. “Toward a Theory of Situation Awareness in Dynamic Systems.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 32–64. doi:10.1518/001872095779049543.

Endsley, M. 2000. “Situation Models: An Avenue to the Modeling of Mental Models.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44 (1): 61–64. doi:10.1177/154193120004400117.

Endsley, M., and D. Garland. 2000. “Pilot Situation Awareness Training in General Aviation.” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 44 (11): 357–360. doi:10.1177/154193120004401107.

Endsley, M., and W. Jones. 2001. “A Model of Inter- and Intrateam Situation Awareness: Implications for Design, Training, and Measurement.” In *New Trends in Cooperative Activities*, edited by M. McNeese, E. Salas, and M. Endsley, 46–67. Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M., and M. Robertson. 2000. “Situation Awareness in Aircraft Maintenance Teams.” *International Journal of Industrial Ergonomics* 26 (2): 301–325. doi:10.1016/S0169-8141(99)00073-6.

Endsley, M., and M. Robertson. 2000. “Training for Situation Awareness.” In *Situation Awareness and Measurement*, edited by M. Endsley and D. Garland, 349–365. Mahwah, NJ: Erlbaum.

Feldon, D., and R. Clark. 2006. “Instructional Implications of Cognitive Task Analysis as a Method for Improving the Accuracy of Experts’ Self-Report.” In *Avoiding Simplicity, Confronting Complexity: Advances in Studying and Designing (Computer-Based) Powerful Learning Environments*, edited by G. Clarebout and J. Elen, 109–116. Rotterdam: Sense Publishers.

Fisher, R., and R. Geiselman. 1992. *Memory Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Charles C Thomas Publisher.

Fisher, R., R. Geiselman, and M. Amador. 1989. "Field Test of the Cognitive Interview: Enhancing the Recollection of Actual Victims and Witnesses of Crime." *The Journal of Applied Psychology* 74 (5): 722–727. doi:10.1037/0021-9010.74.5.722.

Geiselman, R. E., R. P. Fisher, D. P. MacKinnon, and H. L. Holland. 1986. "Enhancement of Eyewitness Memory with the Cognitive Interview." *The American Journal of Psychology* 99 (3): 385–401. doi:10.2307/1422492.

Gilbert, S. 2011. *Models-Based Science Teaching: Understanding and Using Mental Models*. Arlington, VA: NSTA Press.

Goodman, P., E. Ravlin, and M. Schminke. 1987. "Understanding Groups in Organizations." *Research in Organizational Behavior* 9: 121–173.

Grossman, R., S. Friedman, and S. Kalra. 2017. "Teamwork Processes and Emergent States." In *The Wiley Blackwell Handbook of the Psychology of Team Working and Collaborative Processes*, edited by E. Salas, R. Rico, and J. Passmore, 243–269. Hobogen, NJ: Wiley & Sons.

Hackman, J. 1987. "The Design of Work Teams." In *Handbook of Organizational Behavior*, edited by W. Lorsch, 315–342. Englewood cliffs, NJ: Prentice Hall.

Hart, S., and L. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." *Advances in Psychology* 52: 139–183.

Hart, S., and D. McPherson. 1976. "Airline Pilot Time Estimation during Concurrent Activity Including Simulated Flight." In *47th Annual Meeting of the Aerospace Medical Association*, Bal Harbor, FL.

Hodson, D., and R. Hill. 2014. "The Art and Science of Live, Virtual, and Constructive Simulation for Test and Analysis." *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 11 (2): 77–89. doi:10.1177/1548512913506620.

Houck, M., L. Whitaker, and R. Kendall. 1993. *An Information Processing Classification of beyond-Visual-Range Air Intercepts*. Report Number AL/HR-TR-1993-0061. Williams Air Force Base, AZ: Air Force Material Command.

Ilgen, D., J. Hollenbeck, M. Johnson, and D. Jundt. 2005. "Teams in Organizations: From Input-Process-Output Models to IMOI Models." *Annual Review of Psychology* 56: 517–543. doi:10.1146/annurev.psych.56.091103.070250.

Johanssen, G., N. Moray, R. Pew, J. Rasmussen, A. Sanders, and C. Wickens. 1979. "Final Report of Experimental Psychology Group." In *Mental Workload*, edited by N. Moray, 101–114. Boston, MA: Springer.

Johnson, Tristan E., Youngmin Lee, Miyoung Lee, Debra L. O'Connor, Mohammed K. Khalil, and Xiaoxia Huang. 2007. "Measuring Sharedness of Team-Related Knowledge: Design and Validation of a Shared Mental Model Instrument." *Human Resource Development International* 10 (4): 437–454. doi:10.1080/13678860701723802.

Joint Chief of Staff. 2013. *Joint Publication 3-60: Joint Targeting*. Washington, DC.: US Joint Chief of Staff.

Jones, D., and M. Endsley. 1996. "Sources of Situation Awareness Errors in Aviation." *Aviation, Space, and Environmental Medicine* 67 (6): 507–512.

Jonker, C., M. Van Riemsdijk, B. Vermeulen, and F. Den Helder. 2010. "Shared Mental Models: A Conceptual Analysis." In *COIN2010 International Workshops*, edited by M. De Vos, 132–151. Berlin; Springer.

Keeney, R., and H. Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge, UK: Cambridge University Press.

Kelly, M., A. Lee Wooldridge, R. Hennessy, and J. Reed. 1979. "Air Combat Maneuvering Performance Measurement." *Proceedings of the Human Factors Society Annual Meeting* 23 (1): 324–328. doi:10.1177/107118137902300183.

Kerr, N., and R. Tindale. 2004. "Group Performance and Decision Making." *Annual Review of Psychology* 55: 623–655. doi:10.1146/annurev.psych.55.090902.142009.

Klimoski, R., and S. Mohammed. 1994. "Team Mental Model: Construct or Metaphor?" *Journal of Management* 20 (2): 403–437. doi:10.1177/014920639402000206.

Kozlowski, S., S. Gully, E. Nason, and E. Smith. 1999. "Developing Adaptive Teams: A Theory of Compilation and Performance across Levels and Time." In *The Changing Nature of Performance:*

Implications for Staffing, Motivation, and Development, edited by D. Ilgen and E. Pulakos, 240–292. San Fransico, CA: Jossey-Bass.

Kozlowski, S., and D. Ilgen. 2006. “Enhancing the Effectiveness of Work Groups and Teams.” *Psychological Science in the Public Interest: A Journal of the American Psychological Society* 7 (3): 77–124. doi:10.1111/j.1529-1006.2006.00030.x.

Kozlowski, S., & Klein, K. 2000. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In *Multilevel Theory, Research and Methods in Organizations: Foundations, Extensions and New Directions*, edited by K. Klein and S. Kozlowski, 3–90. San Francisco, CA: Jossey-Bass.

Langan-Fox, J., J. Anglim, and J. R. Wilson. 2004. “Mental Models, Team Mental Models, and Performance: Process, Development, and Future Directions.” *Human Factors and Ergonomics in Manufacturing* 14 (4): 331–352. doi:10.1002/hfm.20004.

Langan-Fox, J., S. Code, and K. Langfield-Smith. 2000. “Team Mental Models: Techniques, Methods, and Analytic Approaches.” *Human Factors* 42 (2): 242–271. doi:10.1518/001872000779656534.

Li, W.-C., and D. Harris. 2006. “Pilot Error and Its Relationship with Higher Organizational Levels: HFACS Analysis of 523 Accidents.” *Aviation, Space, and Environmental Medicine* 77 (10): 1056–1061.

Lilja, H., J. Brynielsson, and S. Lindquist. 2016. “Identifying Radio Communication Inefficiency to Improve Air Combat Training Debriefings.” In *Proceedings of the 2016 Interservice/Industry Training, Simulation, and Education Conference*, 1–12. Arlington, VA: National Training and Simulation Association.

Lim, B., and K. Klein. 2006. “Team Mental Models and Team Performance: A Field Study of the Effects of Team Mental Model Similarity and Accuracy.” *Journal of Organizational Behavior* 27 (4): 403–418. doi:10.1002/job.387.

Lutfi, M., and M. Sukkar. 2012. “Effect of Blood Pressure on Heart Rate Variability.” *Khartoum Medical Journal* 4 (1): 548–553.

Mansikka, H., D. Harris, and K. Virtanen. 2019. “Pilot Competencies as Components of a Dynamic Human-Machine System.” *Human Factors and Ergonomics in Manufacturing & Service Industries* 29 (6): 466–477. doi:10.1002/hfm.20809.

Mansikka, H., P. Simola, K. Virtanen, D. Harris, and L. Oksama. 2016. “Fighter pilots’ heart rate, heart rate variation and performance during instrument approaches.” *Ergonomics* 59 (10): 1344–1352. doi:10.1080/00140139.2015.1136699.

Mansikka, H., K. Virtanen, and D. Harris. 2019. “Comparison of NASA-TLX Scale, Modified Cooper-Harper Scale and Mean Inter-Beat Interval as Measures of Pilot Mental Workload during Simulated Flight Tasks.” *Ergonomics* 62 (2): 246–254. doi:10.1080/00140139.2018.1471159.

Mansikka, H., K. Virtanen, and D. Harris. 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. doi:10.1109/THMS.2018.2874186.

Mansikka, H., K. Virtanen, D. Harris, and J. Salomäki. 2019a. “Live–Virtual–Constructive Simulation for Testing and Evaluation of Air Combat Tactics, Techniques, and Procedures, Part 1: Assessment Framework.” *The Journal of Defense Modeling and Simulation*. Ahead of print. doi:10.1177/1548512919886375.

Mansikka, H., K. Virtanen, D. Harris, and J. Salomäki. 2019b. “Live–Virtual–Constructive Simulation for Testing and Evaluation of Air Combat Tactics, Techniques, and Procedures, Part 2: Demonstration of the Framework.” *The Journal of Defense Modeling and Simulation*. Ahead of print. doi:10.1177/1548512919886378.

Marks, M., J. Mathieu, and S. Zaccaro. 2001. “A Temporally Based Framework and Taxonomy of Team Processes.” *The Academy of Management Review* 26 (3): 356–376. doi:10.2307/259182.

Marks, M., S. Zaccaro, and J. Mathieu. 2000. “Performance Implications of Leader Briefings and Team-Interaction Training for Team Adaptation to Novel Environments.” *The Journal of Applied Psychology* 85 (6): 971–986. doi:10.1037/0021-9010.85.6.971.

Mathieu, J., T. Heffner, G. Goodwin, J. Cannon-Bowers, and E. Salas. 2005. “Scaling the Quality of Teammates’ Mental Models: Equifinality and Normative Comparisons.” *Journal of Organizational Behavior* 26 (1): 37–56. doi:10.1002/job.296.

Mathieu, J., M. Maynard, T. Rapp, and L. Gilson. 2008. "Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse into the Future." *Journal of Management* 34 (3): 410–476. doi:10.1177/0149206308316061.

McGrath, J. 1984. *Groups: Interaction and Performance*, Vol. 14. Englewood Cliffs, NJ: Prentice-Hall.

Mohammed, S., R. Klimoski, and J. Rentsch. 2000. "The Measurement of Team Mental Models: We Have No Shared Schema." *Organizational Research Methods* 3 (2): 123–165. doi:10.1177/109442810032001.

Moray, N. 2004. "Ou' sont les neiges d'antan?" Paper Presented at the Human Performance, Situation Awareness and Automation; Current Research and Trends, Orlando, FL, 22–24 March.

Neisser, U. 1976. *Cognition and Reality: Principles and Implications of Cognitive Psychology*. New York: WH Freeman/Times Books/Henry Holt & Co.

O'Donnell, R., Eggemeier, F., & Thomas, F. 1986. Workload assessment methodology. In *Handbook of Perception and Human Performance*, edited by K. Boff and J. Thomas, Vol. 2, pp. 1-42–1-49. New York: Wiley and Sons.

Ohlander, U., J. Alfredson, M. Riveiro, and G. Falkman. 2019. "Fighter Pilots' Teamwork: A Descriptive Study." *Ergonomics* 62 (7): 880–890. doi:10.1080/00140139.2019.1596319.

Paas, F., J. Tuovinen, H. Tabbers, and P. Van Gerven. 2003. "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory." *Educational Psychologist* 38 (1): 63–71. doi:10.1207/S15326985EP3801_8.

Paddon, H. G. 1977. "Maneuvering Target Simulation for Testing the Terminal Guidance of Air-to-Air Missiles." Masters thesis., Air Force Institute of Technology. https://apps.dtic.mil/dtic/tr/fulltext/u2/a039757.pdf.

Parasuraman, R., T. Sheridan, and C. Wickens. 2008. "Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs." *Journal of Cognitive Engineering and Decision Making* 2 (2): 140–160. doi:10.1518/155534308X284417.

Reid, G., and T. Nygren. 1988. "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload." *Advances in Psychology* 52: 185–218.

Roscoe, A., and G. Ellis. 1990. *A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use*. Technical Report No. TR90019. Farnborough: Royal Aerospace Establishment.

Rouse, W., J. Cannon-Bowers, and E. Salas. 1992. "The Role of Mental Models in Team Performance in Complex Systems." *IEEE Transactions on Systems, Man, and Cybernetics* 22 (6): 1296–1308. doi:10.1109/21.199457.

Rouse, W., and N. Morris. 1986. "On Looking into the Black Box: Prospects and Limits in the Search for Mental Models." *Psychological Bulletin* 100 (3): 349–363. doi:10.1037/0033-2909.100.3.349.

Royal Korean Air Force (RoKAF). 2005. *Basic Employment Manual F-16C Koran Air Force Tactics, Techniques and Procedures, 3-3*. Korean Air Force. http://falcon.blu3wolf.com/Docs/Basic-Employment-Manual-F-16C-RoKAF.pdf.

Rumelhart, D. 1984. "Schemata and the Cognitive System." In *Handbook of Social Cognition*, edited by R. Wyer and T. Srull, 161–188. Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D., and A. Ortney. 1977. "The Representation of Knowledge in Memory." In *Schooling and the Acquisition of Knowledge*, edited by R. Anderson, R. Spiro, and W. Montague, 99–135. Hillsdale, NJ: Lawrence Erlbaum.

Salas, E., C. Bowers, S. Cannon-Bowers, and S. Tannenbaum. 1992. "Toward an Understanding of Team Performance and Training." In *Teams: Their Training and Performance*, edited by R. Swezey and E. Salas, 3–29. Norwood, NJ: Ablex.

Salas, E., C. Prince, D. Baker, and L. Shrestha. 1995. "Situation Awareness in Team Performance: Implications for Measurement and Training." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 123–136. doi:10.1518/001872095779049525.

Salas, E., D. Sims, and C. Burke. 2005. "Is There a "Big Five" in Teamwork?" *Small Group Research* 36 (5): 555–599. doi:10.1177/1046496405277134.

Salas, E., K. Stagl, C. Burke, and G. Goodwin. 2007. "Fostering Team Effectiveness in Organizations: Toward an Integrative Theoretical Framework on Team Performance." In *Modeling Complex Systems: Motivation, Cognition and Social Processes: Nebraska Symposium on Motivation*, edited by R. Dienstbier, J. Shuart, W. Spaulding, and J. Poland, Vol. 51, 185–243. Lincholn, NE: University of Nebraska Press.

Salmon, P., N. Stanton, G. Walker, C. Baber, D. Jenkins, R. McMaster, and M. Young. 2008. “What Really is Going On? Review of Situation Awareness Models for Individuals and Teams.” *Theoretical Issues in Ergonomics Science* 9 (4): 297–323. doi:10.1080/14639220701561775.

Salmon, P., N. Stanton, G. Walker, and D. Green. 2006. “Situation Awareness Measurement: A Review of Applicability for C4i Environments.” *Applied Ergonomics* 37 (2): 225–238. doi:10.1016/j.apergo.2005.02.001.

Salmon, P., N. Stanton, G. Walker, D. Jenkins, D. Ladva, L. Rafferty, and M. Young. 2009. “Measuring Situation Awareness in Complex Systems: Comparison of Measures Study.” *International Journal of Industrial Ergonomics* 39 (3): 490–500. doi:10.1016/j.ergon.2008.10.010.

Saner, L., C. Bolstad, C. Gonzalez, and H. Cuevas. 2009. “Measuring and Predicting Shared Situation Awareness in Teams.” *Journal of Cognitive Engineering and Decision Making* 3 (3): 280–308. doi:10.1518/155534309X474497.

Sarter, N., and D. Woods. 1991. “Situation Awareness: A Critical but Ill-Defined Phenomenon.” *The International Journal of Aviation Psychology* 1 (1): 45–57. doi:10.1207/s15327108ijap0101_4.

Schaeffer, N., and S. Presser. 2003. “The Science of Asking Questions.” *Annual Review of Sociology* 29 (1): 65–88. doi:10.1146/annurev.soc.29.110702.110112.

Schank, R., and R. Abelson. 1977. *Scritps, Plans, Goals and Understanding: An Inquiry into Human Knowledge*. Hillsdale, NJ: Lawrence Erlbaum

Schank, R., and R. Abelson. 1988. “Scripts, Plans, Goals, and Understanding.” In *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, edited by A. Collins and E. Smith, 190–223. San Mateo, CA: Morgan Kaufmann.

Selcon, S., T. Hardiman, D. Croft, and M. Endsley. 1996. “A Test-Battery Approach to Cognitive Engineering: To Meta-Measure or Not to Meta-Measure, That is the Question!” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 40 (4): 228–232. doi:10.1177/154193129604000418.

Smith, K., and P. A. Hancock. 1995. “Situation Awareness is Adaptive, Externally Directed Consciousness.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (1): 137–148. doi:10.1518/001872095779049444.

Smith, E., J. Borgvall, and P. Lif. 2007. *Team and Collective Performance Measurement*. Report Number RTO-TR-HFM-121-Part II. Bedfordshire, UK: NATO.

Stanton, N., P. Salmon, L. Rafferty, G. Walker, C. Baber, and D. Jenkins. 2013. *Human Factors Methods: A Practical Guide for Engineering and Design*. Surrey: Ashgate Publishing.

Stanton, N., P. Salmon, G. Walker, and D. Jenkins. 2009. “Genotype and Phenotype Schemata and Their Role in Distributed Situation Awareness in Collaborative Systems.” *Theoretical Issues in Ergonomics Science* 10 (1): 43–68. doi:10.1080/14639220802045199.

Stanton, N., P. Salmon, G. Walker, and D. P. Jenkins. 2010. “Is Situation Awareness All in the Mind?” *Theoretical Issues in Ergonomics Science* 11 (1–2): 29–40. doi:10.1080/14639220903009938.

Stanton, N. A., R. Stewart, D. Harris, R. J. Houghton, C. Baber, R. McMaster, P. Salmon, et al. 2006. “Distributed Situation Awareness in Dynamic Systems: Theoretical Development and Application of an Ergonomics Methodology.” *Ergonomics* 49 (12–13): 1288–1311. doi:10.1080/00140130600612762.

Steiner, I. 1972. *Group Process and Productivity*. New York: Academic Press.

Sterman, J. 2001. “System Dynamics Modeling: Tools for Learning in a Complex World.” *California Management Review* 43 (4): 8–25. doi:10.2307/41166098.

Sulistyawati, K., C. Wickens, and Y. Chui. 2009. “Exploring the Concept of Team Situation Awareness in a Simulated Air Combat Environment.” *Journal of Cognitive Engineering and Decision Making* 3 (4): 309–330. doi:10.1518/155534309X12599553478791.

Sundstrom, E., K. P. De Meuse, and D. Futrell. 1990. “Work Teams: Applications and Effectiveness.” *American Psychologist* 45 (2): 120–133. doi:10.1037/0003-066X.45.2.120.

Svensson, E., M. Angelborg-Thanderez, L. Sjöberg, and S. Olsson. 1997. "Information Complexity-Mental Workload and Performance in Combat Aircraft." *Ergonomics* 40 (3): 362–380. doi:10.1080/001401397188206.

Svensson, E., M. Angelborg-Thanderz, and L. Sjöberg. 1993. "Mission Challenge, Mental Workload and Performance in Military Aviation." *Aviation, Space, and Environmental Medicine* 64 (11): 985–991.

Svensson, E., and G. Wilson. 2002. "Psychological and Psychophysiological Models of Pilot Performance for Systems Development and Mission Evaluation." *The International Journal of Aviation Psychology* 12 (1): 95–110. doi:10.1207/S15327108IJAP1201_8.

Taylor, R. 1990. "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design." In *Situational Awareness*, edited by E. Salas and A. Dietz, 78–111. New York: Ashgate.

Tsang, P., and M. Vidulich. 2006. "Mental Workload and Situation Awareness." In *Handbook of Human Factors and Ergonomics*, edited by G. Salvendy, 243–268. New York: Wiley & Sons.

Van Der Vegt, G. S., and J. S. Bunderson. 2005. "Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification." *Academy of Management Journal* 48 (3): 532–547. doi:10.5465/amj.2005.17407918.

Vidulich, M. 2003. "Mental Workload and Situation Awareness: Essential Concepts for Aviation Psychology Practice." In *Principles and Practice of Aviation Psychology*, edited by P. Tsang and M. Vidulich, 115–146. Mahwah, NJ: Lawrence Erlbaum.

Vidulich, M., and P. Tsang. 2015. "The Confluence of Situation Awareness and Mental Workload for Adaptable Human–Machine Systems." *Journal of Cognitive Engineering and Decision Making* 9 (1): 95–97. doi:10.1177/1555343414554805.

Waag, W., and M. Houck. 1994. "Tools for Assessing Situational Awareness in an Operational Fighter Environment." *Aviation, Space, and Environmental Medicine* 65 (5): 13–19.

Wickens, C. 2002. "Situation Awareness and Workload in Aviation." *Current Directions in Psychological Science* 11 (4): 128–133. doi:10.1111/1467-8721.00184.

Wickens, C., J. Lee, Y. Liu, and S. Becker. 2004. *An Introduction to Human Factors Engineering*. Upper Saddle River, NJ: Pearson Education.

Wierwille, W., and J. Casali. 1983. "A Validated Rating Scale for Global Mental Workload Measurement Applications." *Proceedings of the Human Factors Society Annual Meeting* 27 (2): 129–133. doi:10.1177/154193128302700203.

Willis, G., P. Royston, and D. Bercini. 1991. "The Use of Verbal Report Methods in the Development and Testing of Survey Questionnaires." *Applied Cognitive Psychology* 5 (3): 251–267. doi:10.1002/acp.2350050307.

Wilson, J., and A. Rutherford. 1989. "Mental Models: Theory and Application in Human Factors." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 31 (6): 617–634. doi:10.1177/001872088903100601.

Young, M., K. Brookhuis, C. Wickens, and P. Hancock. 2015. "State of Science: Mental Workload in Ergonomics." *Ergonomics* 58 (1): 1–17. doi:10.1080/00140139.2014.956151.

Zhihua, D., Z. Yuanchang, D. Yanqiang, and M. Xianguo. 2013. "Constructing LVC Simulation Environments Based on Legacy Simulations." In *2013 International Conference on Virtual Reality and Visualization*, 325–328. Xi'an, China: IEEE.