Accuracy and Similarity of Team Situation Awareness in Simulated Air Combat

Heikki Mansikka; Don Harris; Kai Virtanen

BACKGROUND: Fighter pilots' Team Situation Awareness (TSA) has been studied from the perspective of TSA accuracy, which represents how closely the pilots' collective knowledge is aligned with the real world. When TSA accuracy is low, the pilots can have similarly or dissimilarly inaccurate SA. The concept of TSA similarity represents the similarity of team members' collective knowledge. This paper investigates how TSA accuracy and similarity of F/A-18 pilots are associated with performance.

METHOD: Data were extracted from simulated air combat missions. Performance and TSA were investigated in 58 engagements. The accuracy and similarity of pilots' SA were elicited and performance was evaluated. TSA accuracy and similarity were analyzed with respect to the flights' performance, and the independent variables were events in which the flights initiated engagements with enemy aircraft versus events in which the flights were engaged by enemy aircraft.

RESULTS: With the mentioned events as the main effect, there were statistically significant differences at all levels of TSA accuracy and similarity. With performance as the main effect, there were also significant differences at all levels of TSA accuracy and similarity. TSA accuracy and similarity were superior in offensive engagements and when engagements were successful.

DISCUSSION: The main contribution of this paper is the extension of the concept of TSA similarity to air combat: both TSA similarity and accuracy were higher when the flight was engaging the enemy aircraft, compared to situations when the flight itself was being engaged. The results also suggest that low TSA accuracy and similarity have a statistically significantly negative impact on the flights' performance.

KEYWORDS: air combat; team performance output; team situation awareness accuracy; team situation awareness similarity.

Mansikka H, Harris D, Virtanen K. *Accuracy and similarity of team situation awareness in simulated air combat*. Aerosp Med Hum Perform. 2023; 94(6):429–436.

In air combat, pilots must repeatedly make tactical decisions. Ideally, the pilots would possess all available knowledge regarding friendly and enemy entities engaged in air combat. In addition, they would have the time and capacity to scrutinize the potential outcome of each decision alternative before making their final decision. Decisions in air combat, however, must be made with incomplete knowledge and there is seldom time to evaluate completely the expected outcomes of even a few alternatives. While pilots may occasionally reach an acceptable outcome simply by chance, the likelihood of being successful increases if the pilots' decisions are based on adequate situation awareness (SA). The dominant theory of SA describes it as a hierarchical construct with three levels.[8] According to Endsley, SA is the perception of relevant elements in the environment (SA level 1), comprehension of their meaning (SA level 2), and a projection of their status in the near future (SA level 3).

For pilots to make good decisions, they do not need to have knowledge about everything around them. In fact, trying to obtain all knowledge about the entities involved in a fast-paced air combat engagement would be extremely slow and thus counterproductive for SA and decision-making. A person's knowledge is organized in a form of mental representations[30] and stored in and transferred between a long term memory (LTM) and a working memory (WM).[1] When the mental

From the National Defence University, Helsinki, Finland, Aalto University, Helsinki, Finland, and Coventry University, Coventry, United Kingdom.

This manuscript was received for review in October 2022. It was accepted for publication in March 2023.

Address correspondence to: Dr. Heikki Mansikka, Aalto University, P.O. Box 11000, Helsinki FI-00076 AALTO, Finland; heikki.mansikka@aalto.fi.

DOI: https://doi.org/10.3357/AMHP.6196.2023

representations held in LTM are activated (i.e., transferred to WM) and updated with observations, they enable a person to comprehend the situation they are interacting with[17] and to generate descriptions of the situation's observed and predicted states.[25] While the knowledge held in LTM is static and often context-independent, the knowledge in WM is dynamic and context-dependent. These situated and dynamically updated mental representations are referred to as situation models[9] or, more commonly, as SA.[8]

SA is an essential precursor to rational decision-making and response selection, which eventually result in performance outputs.[7] As the outputs modify the environment that a decision-maker is interacting with, the decision-making process essentially forms a reciprocal, cyclical interaction between a decision-maker and the environment.[23] In air combat, both the friendly and enemy pilots have their own decision cycles, often contextualized as Orient, Observe, Decide, Act (OODA) loops.[3] One of the objectives in air combat training is to manipulate the pace of the cognitive and action processes within these loops such that a dominant operational tempo can be achieved.[29]

To gain and maintain a satisfactory SA for timely tactical decision-making, the pilots must allocate their attention and the capacity of their working memory only to the most relevant factors in their tactical environment. An attribute is the smallest relevant unit of which the pilots can have SA; a functional collection of attributes is a concept.[16] For example, a nonfriendly aircraft is a concept whose attributes include position, type, and offensive capabilities, among others.

While there are many different techniques to evaluate pilots' SA, they all essentially attempt to elicit their internal representation of important SA attributes' current and future states. The accuracy of SA is determined by comparing pilots' internally depicted states of the attributes to their states in objective reality, often referred to as "ground truth."[15] The better the pilots' internal representation of the attributes is aligned with the ground truth, the better their SA.

Pilots typically engage in air combat in teams of four (i.e., flights). A flight consists of a flight leader, a two-ship leader, and their wingmen. The flight leader is usually referred to as #1, the two-ship leader as #3, and their wingmen as #2 and #4, respectively. While each pilot has their own internal representation of the flight's task, the flight's tactical environment, and the flight itself,[4] the flight also possesses a shared and organized knowledge regarding these attributes.[19] Terms such as "team mental model," "shared cognition," and "shared knowledge"[12] all essentially refer to this type of common cognitive ground of the flight, each with a slightly different perspective and emphasis. In this paper, the term "Team SA (TSA)" is used to describe the flight members' collective SA.

The relationship between a flight's performance output and TSA is a complex one. For example, a recent study has shown that the relationship is curvilinear, not linear as was previously assumed.[16] In addition, in a highly complex task environment like air combat, the flight's success is not only dependent on the flight's ability to gain and maintain TSA, but also other factors like chance. As a result, equally successful flights can be either good or simply fortunate, and unsuccessful ones can be either unlucky or simply poor. In general, good TSA is an essential contributor to effective decision-making and satisfactory performance output. Therefore, there is a practical need to evaluate flights' TSA in air combat to separate the unfortunate teams from the poor ones.

The flight's TSA has previously been studied almost completely from the perspective of TSA accuracy (TSA ACC).[21] TSA ACC represents how closely the flight's collective knowledge is aligned with the ground truth. For most purposes, the evaluation of TSA ACC is adequate. However, TSA ACC does not draw a complete picture of the nature of the shared knowledge possessed by the flight. For example, the widely used Situation Awareness Global Assessment Technique (SAGAT)[6] estimates TSA based on the overlap of team members' correct responses to SA probes, thus totally ignoring the information about the SA of those team members whose responses are incorrect. If TSA ACC is high, the flight members' SA is closely matched with the ground truth. In such a situation, SA of each flight member is also very similar. If, however, TSA ACC is low, the pilots can have similarly or dissimilarly inaccurate SA. While the idea of TSA similarity (TSA SIM) is not new,[27] two important issues about the flight's collective knowledge in air combat settings are still to be answered. Firstly, previous formal efforts have not investigated how the similarity or dissimilarity of the flight's collective knowledge are associated with its performance output. Secondly, it remains unclear whether team members' similarly or dissimilarly inaccurate SA is better for the flight's performance output. This paper bridges these remaining gaps between TSA as a theoretical construct and air combat as a real-life team activity.

In air combat, the flight's taskwork is what it does, whereas teamwork is about how the flight does it[20] (see the work of Ohlander et al.[22] and Tsifetakis and Kontogianni[28] for more information about teamwork and a recent article by Mansikka et al.[15] about taskwork in an air combat context). Team performance considers the flight's performance output in a given task, whereas team effectiveness takes a more holistic view and takes into account how it achieved it.[11] Both teamwork and taskwork are needed for a flight to be effective[13] as teamwork processes have a mediating role in transforming team members' efforts into the performance output.[18] Finding and intercepting an enemy aircraft are examples of taskwork, whereas teamwork consists of team activities such as closed loop communication, leadership, and development of shared mental models.[26] TSA itself is a product of teamwork.[2] It is an emergent state, which impacts other team- and taskwork processes.[18] This paper concentrates on two critical events related to flights' taskwork: events where the flight engages the enemy aircraft and events where the flight is engaged by the enemy aircraft. An engagement refers to an activity where an aircraft attempts to achieve a weapon firing position against an opposing aircraft, while simultaneously attempting to prevent the opposing aircraft from achieving such a position. This paper explores the mediating role of teamwork in flight's team performance in air combat. More specifically, this paper investigates how TSA ACC and TSA SIM of F/A-18 flights are associated with the flights' success or failure in the events where the flight is either being engaged by the enemy aircraft or vice versa.

METHOD

Subjects

The data were collected during training that the subjects would have undertaken had no experiment existed. In Finland, an ethical review of nonmedical research involving human subjects is based on a set of guidelines drawn up by the Finnish National Board on Research Integrity, TENK. According to the guidelines of TENK, the research configuration of this paper was such that it did not require an ethical review statement from a human sciences ethics committee.

The study was attended by 16 combat-ready F/A-18 pilots. All subjects were men. Their mean flying experience on the F/A-18 was 432 flight hours (SD = 231). All subjects had passed an aeromedical examination within the last 12 mo and were fit to fly. The subjects manned eight simulators to form two friendly flights, each comprising four simulators. The pilots in each flight had flown with each other before the study. A fighter controller was assigned for both flights.

The flight leaders' mean air combat training experience with the F/A-18 was 352 flight hours (SD = 65) whereas the average of the two-ship leaders was 356 flight hours (SD = 183). The mean F/A-18 air combat training experience of the wingmen was 209 flight hours (SD = 150). The F/A-18 air combat training experience of flights varied from 265 to 323 flight hours (SD = 20). The friendly force is hereafter referred to as "Blue" and the enemy force is referred to as "Red".

Equipment

Two types of flight simulators were used in the study: those with manually operated physical controls and others incorporating digital touch screen controls. Both simulators were routinely used in fighter pilot training. Simulators were linked to each other and to the fighter controllers' simulated command and control position via Distributed Interactive Simulation (DIS) connection. Subjects were able to transfer information with each other via datalink and via radio. The systems of both Blue and Red were simulated. The weapon systems were parametrized using probability parameters in order to account for various uncertainties affecting engagements, including probability of detection, probability of guidance, and probability of kill. As in real life, the subjects had to consider these probabilities when making tactical decisions.

Procedure

Data for this study were extracted from air combat missions flown as part of a larger distributed simulator exercise. The exercise included 28 prepared scenarios. Two Blue flights flew in each scenario. Each scenario was repeated twice. Between scenarios, the pilots of the Blue flights were changed such that all 16 subjects were eventually exposed to the same scenarios. As a result, the Blue flights flew a total of 112 missions during the exercise.

Before each scenario, the subjects entered the simulators and the scenario was loaded. The task of Blue varied from scenario to scenario, whereas the task of the computer-generated Red was always to engage Blue. In each scenario, the Blue flights used their standard tactics to complete their tasks, while Red was programmed to replicate threat behavior and tactics briefed for the scenario. The fighter controllers had access to the simulated radar picture of the fighters' operating area. With the help of this radar picture, the fighter controllers assisted their flights with information about the Red entities and other Blue units. Once the simulation was initiated, the scenario was allowed to evolve freely until the training objectives were achieved. The duration of each mission was approximately 40 min.

After each mission, the pilots attended a standard debrief. In the debrief, the pilots had access to their cockpit videos and audio. A computer-animated reconstruction of the mission with all Blue and Red simulation entities was generated to support the mission review. Red Engaged and Blue Engaged events were identified during the debriefs. The Red Engaged event refers to a situation where Blue has launched a weapon against Red. In contrast, the Blue Engaged event refers to a situation where Red has launched a weapon against Blue.

When a Red Engaged or Blue Engaged event was identified, the mission reconstruction and cockpit videos were paused. While paused, the flight identified an SA concept and its attribute which mostly influenced that event's occurrence. These were selected from a list of platform independent concepts and attributes designed for beyond-visual air combat described by Mansikka et al.[16] Once the attribute was identified, the cockpit videos and the mission animation were rewound for 60 s and played again to the point where the engagement occurred. During the replay, the pilots evaluated their SA during that 60-s time period regarding the recognized attribute. The pilots were allowed to pause the replay and zoom the animation in or out while making these evaluations. It was emphasized to the pilots that they should assess their SA as it was during the mission, not as it was during the debrief. It should be emphasized that asking the pilots to evaluate their SA during a previously flown mission with the ground truth simultaneously visible is a standard protocol in air combat debriefs.

Once the replay reached the Red Engaged or Blue Engaged event, the debrief was again paused. During this pause, the pilots first scored their SA accuracy regarding the attribute in question. SA accuracy was scored against SA levels 1, 2, and 3 following the protocol described in Mansikka et al.[16] Scoring was conducted by asking the pilots a question related to each SA level. They selected the most appropriate answer for each question from the list of answers. Each answer alternative was associated with a corresponding SA accuracy score ranging from 1 (most inaccurate) to 3 (most accurate). Use of a simple three-point scale enhanced the reliability of SA accuracy and similarity assessments during the debriefs.[14] The SA accuracy questions, answer options, and SA scores are summarized in **Table I**.

Table I. SA Accuracy Questions, Answer Options, and SA Accuracy Scores.

SA LEVEL	SA ACCURACY QUESTION	ANSWER OPTION AND THEIR CORRESPONDING SA ACCURACY SCORES		
		SCORE 1	SCORE 2	SCORE 3
1	How was your perception regarding the attribute?	Inaccurate. The inaccuracies had a significant negative impact on my tactical decision-making.	Inaccurate. The inaccuracies had no significant negative impact on my tactical decision-making.	Accurate or almost accurate. Possible slight inaccuracies had no significant negative impact whatsoever on my tactical decision-making.
2	How was your understanding regarding the attribute's tactical meaning?			
3	How was your anticipation regarding the attribute's state in the near future?			

The pilots tend to assess the accuracy of their SA against the impact it has on their decision-making. The verbal descriptions of the SA accuracy scores in Table I were worded to support this tendency such that they differentiate the significant negative impact, no significant negative impact, and no negative impact options.

After the SA accuracy scores were acquired, the pilots evaluated the similarity of their SA to that of other flight members. The scoring of similarity was conducted such that first #1 was encouraged to elicit his level 1 SA. The SA elicitation was assisted with the SA elicitation questions presented in **Table II**. Once #1 had verbalized his level 1 SA, #2, #3, and #4 compared their level 1 SA to the verbally expressed SA of #1 and selected the most fitting option from SA similarity alternatives provided in Table II. Each option for similarity was associated with an SA similarity score, which ranged from 1 (most dissimilar) to 3 (most similar). Similar to the scoring of accuracy, the verbal descriptions of the SA similarity scores in Table II were worded such that they would differentiate the significant negative impact, no significant negative impact, and no negative impact options. Secondly, #2 expressed his level 1 SA, while #3 and #4 compared the similarity of their SA to that of #2 in the same fashion as was done in the first step. Thirdly, #3 stated his level 1 SA and #4 evaluated the similarity of his SA compared to that of #3. This procedure was repeated for all SA levels such that the SA similarity of all dyads was obtained.

Once all pilots' individual SA accuracy and similarity scores were obtained, the performance output of the event was determined either as Failure or Success. The Red Engaged event, i.e., a situation where Blue had launched a weapon against Red, can result in two possible performance outputs: the enemy aircraft can either be hit or it can survive. From the flight's perspective, the former is considered Success and the latter is Failure. In comparison, the Blue Engaged event, i.e., a situation where Red had launched a weapon against Blue, has opposite goals and success criteria. In terms of the performance output, a flight member being hit is Failure, and the flight being able to evade is considered Success.

When the individual SA accuracy and SA similarity had been scored and the flight's performance output in the event had been logged, the debrief was continued until the next Blue Engaged or Red Engaged event was identified. At that point, the previously described SA accuracy and similarity scoring as well as the determination of the flight's output performance in the event of interest were repeated. SA accuracy and similarity scoring as well as performance evaluation were limited to critical events, which occurred before the first Blue flight member was killed. The rationale was that after that point, the flight was no longer complete and would have different dynamics than that of a flight of four pilots. In sorties where none of the Blue flight members were killed during the mission, the procedure was repeated until the first Red aircraft was killed. Based on this logic, 58 critical events were included in the analysis: 29 Blue Engaged events and 29 Red Engaged events.

After the debrief, the flight's SA level 1-3 TSA ACC scores for an event were determined by calculating the average of flight members' individual SA accuracy scores in respective SA levels. SA level 1-3 TSA SIM scores were determined in the same fashion by calculating the average of dyads' SA level 1-3 similarity scores for an event. As a result, the flights' TSA SIM and TSA ACC scores ranged from 1 to 3.

RESULTS

TSA ACC and TSA SIM scores were analyzed with respect to flights' performance (i.e., Failure/Success in critical events) and Red/Blue Engaged as independent variables. The unit of analysis was at flight level, not for each individual pilot.

To minimize the family-wise probability of a type I error, SA level 1-3 TSA ACC and TSA SIM data as dependent variables were subject to a single Multivariate Analysis of Variance

Table II. SA Elicitation Questions, SA Similarity Alternatives, and SA Similarity Scores.

SA LEVEL	SA ELICITATION QUESTION	SA SIMILARITY ALTERNATIVES AND THEIR SA SIMILARITY SCORES		
		SCORE 1	SCORE 2	SCORE 3
1	Briefly describe how was your perception regarding the attribute?	Dissimilar. The dissimilarity had a significant negative impact on the flight's tactical decision making.	Dissimilar. The dissimilarity had no significant negative impact on the flight's tactical decision making.	Similar or almost similar. Possible slight dissimilarity had no significant negative impact whatsoever on the flight's tactical decision making.
2	Briefly describe how did you comprehend the tactical meaning of the attribute?			
3	Briefly describe how did you expect the attribute's short term status to change?			

(MANOVA) with main effects of Blue Engaged/Red Engaged events and Success/Failure performance outputs. This also removed the effects of any intercorrelation between the dependent variables on the main effects. All estimates of observed power are based upon an alpha level of 0.05.

Both main effects were statistically significant. There was an overall difference in TSA accuracy and TSA similarity for the Blue Engaged and Red Engaged events (Wilks' Lambda = 0.407; $F_{(6,49)} = 11.918$; $P < 0.001$; partial eta^2 = 0.593; observed power = 1.000) and for the Success and Failure performance outputs (Wilks' Lambda = 0.717; $F_{(6,49)} = 3.231$; $P < 0.01$; partial eta^2 = 0.283; observed power = 0.893). The interaction term was nonsignificant (Wilks' Lambda = 0.812; $F_{(6,49)} = 1.886$; $P > 0.05$; partial eta^2 = 0.188). To aid the interpretation of the multivariate results, the significant main effects were further analyzed using univariate factorial analyses of variance (ANOVA) with each TSA accuracy and similarity level as a dependent variable.

With the Blue Engaged and Red Engaged events as the main effect, there were statistically significant differences at all three SA levels of TSA ACC. At SA level 1 TSA ACC, $F_{(1,54)} = 58.749$, $P = 0.000$, partial eta^2 = 0.520, and observed power = 1.000. At SA level 2 TSA ACC, $F_{(1,54)} = 44.606$, $P < 0.001$, partial eta^2 = 0.452, and observed power = 1.000. At SA level 3 TSA ACC, $F_{(1,54)} = 37.534$, $P < 0.001$, partial eta^2 = 0.410, and observed power = 1.000. In all cases, TSA ACC was significantly superior when friendly forces were engaging the enemy (i.e., in Red Engaged events) than when the friendly forces were being attacked (see **Table III**). With TSA SIM as the dependent variable, again, all three univariate ANOVAs were significant. At SA level 1 TSA SIM, $F_{(1,54)} = 43.301$, $P < 0.001$, partial eta^2 = 0.445, and observed power = 1.000. At SA level 2 TSA SIM, $F_{(1,54)} = 39.604$, $P = <0.001$, partial eta^2 = 0.423, and observed power = 1.000. At SA level 3 TSA SIM, $F_{(1,54)} = 33.987$, $P < 0.001$, partial eta^2 = 0.386, and observed power = 1.000. TSA SIM at all three SA levels was better when Blue was engaging Red than when Red was engaging Blue (see Table III).

With the Success and Failure performance outputs as the main effect, there were also statistically significant differences at all three SA levels of TSA ACC. At SA level 1 TSA ACC, $F_{(1,54)} = 8.647$, $P < 0.01$, partial eta^2 = 0.138, and observed power = 0.823. At SA level 2 TSA ACC, $F_{(1,54)} = 5.340$, $P < 0.05$, partial eta^2 = 0.090, and observed power = 0.622. At SA level 3 TSA ACC, $F_{(1,54)} = 14.684$, $P < 0.001$, partial eta^2 = 0.214, and observed power = 0.964. In all cases, TSA ACC was significantly superior when the performance was Success compared to Failure (see Table III). With TSA SIM as the dependent variables, again, all three univariate ANOVAs were statistically significant (SA level 1 TSA SIM: $F_{(1,54)} = 4.869$, $P < 0.05$, partial eta^2 = 0.083, observed power = 0.582; SA level 2 TSA SIM: $F_{(1,54)} = 4.717$, $P < 0.05$, partial eta^2 = 0.080, observed power = 0.569; SA level 3 TSA SIM: $F_{(1,54)} = 10.564$, $P < 0.01$, partial eta^2 = 0.164, observed power = 0.891). TSA SIM was always better when the performance output of a flight was Success rather than Failure.

DISCUSSION

TSA is a product of teamwork and an essential contributor for the flights' decision-making and performance output in air combat. Previous studies of TSA have mainly concentrated on TSA ACC[21] and to the authors' knowledge SA SIM has not been investigated in air combat context before now. In this paper, both TSA ACC and TSA SIM were analyzed. When TSA ACC is low, the team members' SA could be either similarly or dissimilarly inaccurate. It is reasonable to assume that when the team members' SA ACC is low, as often is the case in complex and dynamic environments, it is quite likely that their SA is at least to some extent dissimilar. Therefore, to better understand

Table III. Means (M), Standard Deviations (SD), and Sample Sizes (*N*) Of SA Level 1-3 TSA ACC and TSA SIM Scores for the Blue Engaged and Red Engaged Events and the Performance Outputs, i.e., Failure and Success of those Events.

		BLUE ENGAGED			RED ENGAGED			TOTAL		
		M	SD	*N*	M	SD	*N*	M	SD	*N*
SA level 1 TSA ACC	Failure	1.40	0.48	15	2.73	0.58	14	2.04	0.85	29
	Success	2.04	0.85	14	2.97	0.13	15	2.52	0.75	29
	Total	1.71	0.74	29	2.85	0.42	29	2.28	0.83	58
SA level 1 TSA SIM	Failure	1.53	0.68	15	2.62	0.61	14	2.06	0.84	29
	Success	1.92	0.79	14	2.94	0.22	15	2.45	0.77	29
	Total	1.72	0.75	29	2.79	0.48	29	2.25	0.82	58
SA level 2 TSA ACC	Failure	1.53	0.65	15	2.73	0.58	14	2.11	0.86	29
	Success	2.04	0.85	14	2.97	0.13	15	2.52	0.75	29
	Total	1.78	0.78	29	2.85	0.42	29	2.31	0.83	58
SA level 2 TSA SIM	Failure	1.56	0.72	15	2.61	0.61	14	2.06	0.85	29
	Success	1.93	0.80	14	2.94	0.22	15	2.46	0.76	29
	Total	1.74	0.77	29	2.78	0.47	29	2.26	0.82	58
SA level 3 TSA ACC	Failure	1.40	0.57	15	2.39	0.60	14	1.88	0.77	29
	Success	2.04	0.85	14	2.95	0.14	15	2.51	0.75	29
	Total	1.71	0.78	29	2.68	0.51	29	2.19	0.82	58
SA level 3 TSA SIM	Failure	1.44	0.69	15	2.33	0.60	14	1.87	0.78	29
	Success	1.92	0.79	14	2.91	0.24	15	2.43	0.76	29
	Total	1.67	0.77	29	2.63	0.53	29	2.15	0.81	58

the shared knowledge possessed by a flight, it is necessary to measure its SA SIM together with SA ACC.

While high TSA ACC and TSA SIM are not necessarily the flights' primary objectives or end-states, these emergent states have a role in the flights' success.[18] High TSA ACC and TSA SIM contribute to the flights' performance output by enabling effective coordination of the flight members' activities. Coordination can be either explicit or implicit.[10] With high TSA ACC and TSA SIM, the flight members are able to anticipate each other's actions and information needs without having to communicate the activities regarding decision-making.[24] As such, implicit coordination enables the team to rapidly synchronize its members' activities. If TSA SIM is low, the flight has to rely on communication to explicitly coordinate its work. Situations where both TSA ACC and TSA SIM are low can be highly confusing to the team and may require excessive communication to enable coordinated action. Situations where TSA ACC is low and TSA SIM is high can cause different issues as the team may not even recognize the need for explicit coordination. In both cases, the teams' performance output is likely to fall short of its full potential. Communication as a means of coordination in air combat is slow and vulnerable to deception and jamming. In summary, high TSA ACC and TSA SIM have an essential mediating role in enabling a flight's critical coordination functions and eventually leading to a team's success in air combat (see Table III).

The tasks of engaging an aircraft and being engaged by aircraft have fundamentally conflicting objectives and involve different piloting activities. However, purely from the perspective of pilots' technical and mental demands, the difference between these tasks is unclear. To the authors' knowledge, the existing literature does not present evidence of one type of task being more demanding than the other one. The results of this study clearly indicate that both TSA ACC and TSA SIM were higher when the flight was engaging the Red aircraft, compared to situations when the flight itself was being engaged. It can be rationalized that high accuracy and similarity enabled the flights to grasp the initiative and reach such a tactical advantage that they were able to engage Red. However, not even the highest TSA ACC and TSA SIM can guarantee success in every Red Engaged situation. For example, after Blue had released its weapon at desired launch parameters, Red sometimes performed aggressive evasive maneuvers which resulted in Blue Failure—or Red Success, depending on the perspective. Weapons such as modern air-to-air missiles are technical systems which hit their target with a certain probability affected by several factors. Pilots can manage those probabilities only to a certain extent. In contrast, it is likely that the combination of low TSA ACC and low TSA SIM created coordination problems within the flight which eventually resulted in a flight member being engaged. Even then, reasonable accuracy and similarity enabled the flight to recover from the adverse situation. As the results indicate, in the Blue Engaged situations which resulted in Failure, TSA ACC and TSA SIM were lowest compared to situations which led to Success. The results support an argument that as the flights had reached a tactical advantage, they were in a better position to control the situation and it was relatively easy for them to maintain control once they gained high TSA. In comparison, when the flights had already lost the advantage, there was an increased risk of them becoming reactive. Once that happened, regaining lost TSA became difficult, effective decision-making suffered, and the flights' performance output deteriorated.

The results of the individual univariate ANOVAs decomposing the main effects show that when Red/Blue Engaged was the independent variable, TSA ACC and SIM at level 1 accounted for the most variance, followed by TSA ACC and SIM at levels 2 and 3. This complements the results observed by Mansikka et al.,[16] which suggested that the relationship between TSA and performance was curvilinear. When engagement Success/Failure was the main effect, though, TSA ACC and SIM at level 1 still accounted for more variance than level 2, but the greatest variance in a successful engagement was TSA ACC and SIM at level 3, i.e., the ability to project ahead.[8]

All statistically significant results suggested a substantial size of effect, with the vast majority of them having partial eta-squared values of well in excess of 0.14, which is considered to be large.[5] The significant main effects accounted for 28–59% of the experimental variance, suggesting that the results had substantial functional differences. The variance accounted for in TSA ACC and SIM at each level of TSA was substantially smaller, accounting for 8–16%. However, the majority of significant results had an observed power greater than 0.8, suggesting a type II error probability of below 0.2 in most cases.[5]

While the concept of TSA SIM has been discussed in theory and investigated in several domains, it has not, until now, been applied to air combat. While it is logical that low TSA SIM can only appear together with low TSA ACC, the results of this study clearly indicate that low TSA SIM has a statistically significantly negative impact on the flights' performance output (see Table III). The situation becomes more complex if low or high TSA SIM are accompanied by low TSA ACC. Answering these questions opens interesting possibilities for future research.

The overall approach and findings of this paper should be useful for those responsible for the administration of air combat training curricula. If low TSA ACC and TSA SIM repeatedly occur in a certain training situation among pilots undergoing a similar training program, the observation can be used to detect latent issues in the training curriculum. In addition, if the pilots frequently face training situations where both TSA ACC and TSA SIM are low, the finding should motivate the training organization to look for fundamental errors of omission in the contents of the pilots' training curriculum. Also, situations resulting repetitively in low TSA ACC and high TSA SIM can reveal training items accidently left out of the training curriculum. It should be noted, however, that like any retrospective elicitation technique, the one introduced in this paper is subject to pilots' recall bias. The possibility of pilots untruthfully reporting a high SA cannot be ruled out either. While common access to the ground truth and a healthy organizational culture are likely to reduce such biases, the technique presented in this paper is only as good as the pilots' ability to recall and willingness to verbalize their cognitive processes and statuses during the mission that is being debriefed. Despite the potential challenges, even an attempt to apply the technique in air combat training is valuable as it focuses the flight's attention to events and activities relevant to TSA, thereby adding value simply by promoting reflection and constructive debate.

Most SA measuring techniques suitable for an operational air combat training context are somewhat time-consuming and labor-heavy. Further studies are needed to explore ways to minimize the time and effort it takes to elicit or otherwise disclose pilots' SA. In addition, further studies are needed to investigate if the findings of this paper can be generalized to other aviation domains besides air combat. For the time being, the question of whether team members' similarly or dissimilarly inaccurate SA is better for the teams' performance output in domains other than air combat remains unanswered.

The flight leaders are generally more experienced than two-ship leaders and wingmen are usually less experienced than two-ship leaders. In training setups, such as in missions used for this study, the pattern may not always be as clear. For example, a two-ship leader may be an experienced flight leader supervising a novice flight leader still under training. The pilots' abilities to gain and maintain SA are likely to reflect their experience differences. As the objective of this study was to examine TSA SIM and TSA ACC in a natural setting, no attempt was made to minimize the pilots' experience differences within the flights. Also, the approach to the assessment of TSA SIM and ACC was not intended to direct the training per se. Instead, it was purposed to highlight the importance of the assessment of both TSA SIM and TSA ACC when evaluating the effectiveness of training. The process by which both forms of TSA were calculated involved averaging the scores across the flight. This may diminish the effect of individual SA differences of pilots. Furthermore, not all members of the flight may contribute equally to TSA. The contribution of the flight leader may be disproportionate to TSA and hence also to performance. This should be further investigated in future studies. This paper, however, provides a sound starting point for such explorations.

ACKNOWLEDGEMENTS

Financial Disclosure Statement: The authors have no competing interests to declare.

Authors and Affiliations: Heikki Mansikkia, Ph.D., M.A., Department of Military Technology, National Defence University, Helsinki, Finland; Don Harris, Ph.D., B.Sc., Faculty of Engineering, Environment, and Computing, Coventry University, Coventry, United Kingdom; and Kai Virtanen, Dr. (tech), M.Sc., Aalto University, Helsinki, Finland.

REFERENCES

1. Atkinson R, Shiffrin R. Human memory: a proposed system and its control processes. In: Spence K, editor. The psychology of learning and motivation: advances in research and theory. New York (NY): Academic Press; 1968:89–195.
2. Bolstad C, Endsley M. Measuring shared and team situation awareness in the army's future objective force. Proc Hum Factors Ergon Soc Annu Meet. 2003; 47(3):369–373.
3. Bryant DJ. Rethinking OODA: toward a modern cognitive framework of command decision making. Mil Psychol. 2006; 18(3):183–206.
4. Cannon-Bowers JA, Salas E, Converse S. Shared mental models in expert team decision making. In: Castellan NJ, editor. Individual and group decision making: current issues. New York (NY): Lawrence Erlbaum Associates; 1993:221–246.
5. Cohen J. Statistical power analysis for the behavioral sciences. 2nd rev. ed. Hillsdale (NJ): Erlbaum; 1988.
6. Endsley M. Situation awareness global assessment technique (SAGAT). Proceedings of the IEEE 1988 National Aerospace and Electronics Conference; July 17, 1988; Dayton, OH. Dayton (OH): IEEE; 1988.
7. Endsley M. A survey of situation awareness requirements in air-to-air combat fighters. Int J Aviat Psychol. 1993; 3(2):157–168.
8. Endsley M. Toward a theory of situation awareness in dynamic systems. Hum Factors. 1995; 37(1):32–64.
9. Endsley M. Situation models: an avenue to the modeling of mental models. Proc Hum Factors Ergon Soc Annu Meet. 2000; 44(1):61–64.
10. Espinosa JA, Lerch FJ, Kraut RE. Explicit versus implicit coordination mechanisms and task dependencies: one size does not fit all. In: Salas E, Fiore SM, editors. Team cognition: understanding the factors that drive process and performance. Washington (DC): American Psychological Association; 2004:107–129.
11. Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: from input-process-output models to IMOI models. Annu Rev Psychol. 2005; 56(1):517–543.
12. Klimoski R, Mohammed S. Team mental model: construct or metaphor? J Manage. 1994; 20(2):403–437.
13. LePine JA, Piccolo RF, Jackson CL, Mathieu JE, Saul JR. A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. Pers Psychol. 2008; 61(2): 273–307.
14. Louangrath P. Reliability and validity of survey scales. Int J Res Methodol Soc Sci. 2018; 4(1):50–62.
15. Mansikka H, Virtanen K, Harris D, Jalava M. Measurement of team performance in air combat – have we been underperforming? Theor Issues Ergon Sci. 2021; 22(3):338–359.
16. Mansikka H, Virtanen K, Uggeldahl V, Harris D. Team situation awareness accuracy measurement technique for simulated air combat – curvilinear relationship between awareness and performance. Appl Ergon. 2021; 96:103473.
17. Markman AB, Gentner D. Thinking. Annu Rev Psychol. 2001; 52(1): 223–247.
18. Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team process. Acad Manage Rev. 2001; 26(3):356–376.
19. Mathieu JE, Heffner TS, Goodwin GF, Salas E, Cannon-Bowers JA. The influence of shared mental models on team process and performance. J Appl Psychol. 2000; 85(2):273–283.
20. Mathieu J, Maynard M, Rapp T, Gilson L. Team effectiveness 1997-2007: a review of recent advancements and a glimpse into the future. J Manage. 2008; 34(3):410–476.
21. Nguyen T, Lim CP, Nguyen ND, Gordon-Brown L, Nahavandi S. A review of situation awareness assessment approaches in aviation environments. IEEE Syst J. 2019; 13(3):3590–3603.
22. Ohlander U, Alfredson J, Riveiro M, Falkman G. Fighter pilots' teamwork: a descriptive study. Ergonomics. 2019; 62(7):880–890.
23. Plant KL, Stanton NA. What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit. Ergonomics. 2013; 56(8):1232–1250.
24. Rico R, Sánchez-Manzanares M, Gil F, Gibson C. Team implicit coordination processes: a team knowledge–based approach. Acad Manage Rev. 2008; 33(1):163–184.

25. Rouse WB, Morris NM. On looking into the black box: prospects and limits in the search for mental models. Psychol Bull. 1986; 100(3): 349–363.
26. Salas E, Sims DE, Burke CS. Is there a “big five” in teamwork? Small Group Res. 2005; 36(5):555–599.
27. Sætrevik B, Eid J. The “similarity index” as an indicator of shared mental models and situation awareness in field studies. J Cogn Eng Decis Mak. 2014; 8(2):119–136.
28. Tsifetakis E, Kontogiannis T. Evaluating non-technical skills and mission essential competencies of pilots in military aviation environments. Ergonomics. 2019; 62(2):204–218.
29. US Air Force. Air Force Doctrine Document 1. 2003. [Accessed January 26, 2023]. Available from https://www.doctrine.af.mil/Portals/61/documents/AFDP_1/AFDP-1.pdf.
30. Wilson JR, Rutherford A. Mental models: theory and application in human factors. Hum Factors. 1989; 31(6):617–634.