Improving the State of the Art for Training Human-AI Teams

Technical Report #6B

A New Model for Teaming

July 2026

Lillian Asiala* and Ranjani Narayanan^

****Sonalysts, Inc.***

^Georgia Tech, School of Aerospace Engineering

This Page Intentionally Blank

Table of Contents

This Page Intentionally Blank

List of Acronyms

This Page Intentionally Blank

Abstract

Sonalysts, Inc. (Sonalysts) is working on an initiative to expand our expertise in teaming to include Human-Artificial Intelligence (AI) teams. This effort has centered on the design of an open-source testbed to support hybrid team research. In parallel, the team has also reviewed literature and conversed with experts in the field to understand the theoretical underpinnings for the design and implementation of intelligent agents that could function effectively as teammates. The current report is one in a series of technical reports that describe this foundational research on teaming, and by extension, hybrid teaming. In the previous report in this series, we explored the concept of mental models at the individual and team levels. In the current report, we delve into prominent existing theories of mental models and their implications for teaming. We propose an extension of teaming research and summarize relevant prospective data collection tools for this work.

This Page Intentionally Blank

1 Introduction

When people confront complex, goal-driven problems, they often rely on an *internal simulation* of the situation—a mental representation that serves as a model of the problem and its context (Johnson-Laird, 1980). For a team to succeed on such problems, the team members' mental models must exhibit a degree of continuity. Scholars, however, remain divided on what that continuity should look like: does effective teaming require substantial overlap between teammates' models, or can complementary, non-overlapping models also support coordinated performance? In this report, we (1) review the leading theoretical models of teaming, (2) propose an extension of an existing framework to broaden research on human-AI hybrid teams, and (3) outline a preliminary research plan that identifies useful experimental paradigms and poses initial research questions for the proposed extension.

In our previous technical report[1], we outlined the theoretical foundations of mental models and examined several conceptual frameworks for team interdependencies and workflows. One such framework included Terrien and colleagues' notion of "collective coordination" (Terrien et al., 2020). Collective coordination rests on the idea that team members have shared knowledge about the task, the equipment, and the environmental factors that influence performance. Terrien et al. highlight the role of communication in achieving coordination and noted that overlapping knowledge can improve team performance, but collective coordination does not go as far as to examine the *contents* of individual mental models on a moment-to-moment basis. Consequently, their framework does not make a claim one way or another to distinguish between shared versus complementary mental model frameworks for teaming. To date, no research paradigm captures the fine-grained, episode-level processing of teamwork that tracks each participant's mental model in real time. Existing studies of team mental models remain at the level of static knowledge representations rather than dynamic, intra-individual processing.

Some scholars argue that measuring individual mental representations offers little insight into team performance. For example, in Cooke et al.'s 2013 article on teaming, the authors describe the theory of shared cognition as one based on an information processing model of individual cognition. Under this paradigm, the information processing framework used to study individual cognition is extended to team cognition by treating shared knowledge of team members as the input to team processes such as communication and other coordinating behaviors, which in turn impact team outcomes. Cook et al. (2013) suggest that an information-processing account of team cognition is insufficient for explaining team cognition. This view is, at least in part, based on an assumption that mental models are captured as static snapshots of knowledge rather than dynamic, miniature simulations that support encoding, decision-making, and rapid response. In their review, Cooke et al. (2013) state "the focus on knowledge structure within and across team members, rather than interaction processes, methodologically restricts the Shared Cognition perspective to the individual as the unit of analysis." But that "there is no straightforward mapping between individual team members and the team in action." They also state that "the focus on individuals' cognitive structure lends itself to dealing with task dynamics in terms of static snapshots (often retrospectively reported), missing much of the actual team cognition that unfurls only over time as the situation unfolds." Thus, the issue with the individual as a unit of analysis seems less about the locus of team cognition occurring at the individual level theoretically and more about the limitations of measuring an individual's mental model as a situation unfolds over time in the context of the team and the environment. However, other researchers have found ways to measure ongoing

[1] A copy of this technical report can be provided upon request.

mental representations at the individual level. For instance, Endsley's 2000 article on situation awareness highlights the concept of the situation model - a continuously updated mental representation that changes over time as it tracks changes in the environment. Because situation awareness is tightly linked to effective decision-making in fast-moving tasks, investigating how individual situation models evolve under different teaming conditions is a valuable research direction. A dynamic, information-processing-based framework would (1) improve our ability to predict cognitive activity within teams and (2) inform the design of AI teammates that can foster, adapt, and maintain appropriate mental models in ever-changing collaborative settings. This report outlines the next steps needed to develop such a framework for team environments.

This Page Intentionally Blank

2 Existing Frameworks

Two influential frameworks dominate the study of cognition in team settings: that of Shared Mental Model (SMM) or “shared cognition”, and Interactive Team Cognition (ITC). An essential difference between these two theories is the locus of cognitive activity that drives team performance. The SMM framework treats team performance as a product of the similarity or complementary nature of mental models among team members. Researchers typically measure individuals’ models, compare them across the team, and infer that greater convergence predicts better performance. In this view, the primary lever for improving teamwork is increasing overlapping or complementary knowledge. By contrast, ITC focuses on the processes of communication and coordination. Cognition is regarded as an emergent property of the team, with individuals serving as components of a collective cognitive system. Under ITC, successful performance depends on how well team members coordinate their actions, not on how coordinated team members’ mental models are with one another. Substantial differences, or even incompatibilities among individual models, are therefore permissible. The following sections provide detailed descriptions of each theory, accompanied by illustrative examples from empirical literature.

2.1 Shared Mental Models

Individuals do not perceive the world directly. Instead, they tend to simulate a subset of environmental details internally through the construction of a mental model (Johnson-Laird, 1980). These models allow people to predict and explain their surroundings (Allen, 1997). The formation and manipulation of mental models is the cornerstone of deductive reasoning (Johnson-Laird, 2005) and has implications for situation awareness and assessment (Endsley, 2000). The SMM framework adopts an Input-Process-Outcome (IPO) paradigm to study how team activity unfolds. The inputs consist of everything that exists before a given performance episode begins: each member’s knowledge, skills, and personal attributes; the way the team is composed and organized; the broader organizational context; and the degree to which teammates already possess similar or complementary mental models. A performance episode is a clearly bounded interval of time in which the team carries out the task, receives feedback on its actions, and attempts to translate those pre-existing inputs into coordinated behavior. During this episode, the processes are the concrete actions that convert inputs into results—behaviors such as communicating, forming a joint strategy, coordinating roles and timing, and cooperating to achieve the task. The outcomes are the valued results that emerge from the episode, including task accuracy, speed, safety, mission success, or any other performance metric the team deems important. In this framework, the nature of the individual team member's mental models and how well they facilitate team processes determines performance within each episode (Mathieu et al, 2000).

In the SMM framework, the mental models of team members who work successfully as a team are highly compatible. Researchers describe two types of mental model synergy among team members that result in compatibility. The first is “mental model overlap” – the extent to which the mental models held by different team members are alike. When overlap is high, teammates tend to exhibit comparable cognitive behaviors, making coordination easier because everyone operates from a common mental foundation. The second form of synergy is mental-model complementarity, also called “distributed knowledge”. Here, the individual models may differ markedly, yet each teammate possesses the information that others lack, so that, collectively, the team has a complete picture of the task environment. The advantages of these synergies are dependent on the task and the performance environment. In some performance environments, it could be most advantageous to have a high degree of overlap in knowledge about the task, equipment, and teammate roles, abilities, and expectations. In other situations, compatibility may be characterized by some critical information overlap, complemented by relevant expertise that resides in specific team members. Whether mental models contain compatible or similar information, everyone has a common “mental blueprint” that supports smooth coordination and rapid adaptation to task demands and one another’s actions (Cannon-Bowers & Salas, 2001). The degree of compatibility of mental models influences team processes like communication, strategy, and cooperation, influencing team performance (Mathieu et al, 2000). SMMs also facilitate implicit communication among teammates, leading to better adaptability, especially in scenarios where explicit communication is inhibited (Cannon-Bowers, Salas, & Converse, 1993).

Complementarity in mental models is theoretically linked to better team performance, although there is a significant gap in the literature on this topic (DeChurch & Mesmer-Magnus, 2010). The literature that has explored the complementarity of mental models in teaming has done so through the concept of transactive memory. Building on Wegner’s work (1987, 1995), Mohammed and Dumville (2001) describe transactive memory as a network of individual memory systems in which members not only store their own knowledge but also share awareness of what each other knows. Imagine a computer network: every node (team member) in the network has its own directory of information and a directory of the other nodes’ contents. Because each person knows who holds which piece of expertise, the team can route the right information to the right person quickly and efficiently. This specialization reduces the cognitive load on any single member, expands the overall expertise available to the group, and minimizes unnecessary redundancy (Mohammed & Dumville, 2001). Yoo and Kanawattanachai (2001) conducted a study on undergraduate student teams in a business simulation where they had to make decisions related to marketing, production, and finance to maximize the simulated firm’s stock price. In this study, the authors operationalized the team’s transactive memory system as shared knowledge about who knew what on the team, a construct measured through agreement on survey items such as “The team has a good map of each other’s talents and skills.” They also measured “collective mind”, which was defined as the team’s ability to interrelate actions and decisions with one another. Transactive memory formed early in the simulation, which appeared to foster a stronger collective mind among the team during the project. Both transactive memory and collective mind were related positively to team performance in the simulation. Gupta and Hollingshead (2010) defined “differentiated” and “integrated” transactive memory, two constructs that map well to “mental model complementarity” and “mental model overlap,” respectively. Differentiated transactive memory is where each team member holds distinct information, while integrated transactive memory involves team members with the same knowledge shared across the team. Their study found that groups with differentiated transactive memory solved tasks faster and more accurately than teams that did not. However, the utility of differentiation in teammate knowledge was moderated by the task – teams with strong differentiated transactive memory systems performed better on dynamic and volatile tasks, but did not show the same performance improvement for stable tasks (Gupta & Hollingshead, 2010).

The bulk of published literature on shared mental models has measured “mental model overlap” in teams. In a review of literature on shared mental models conducted by DeChurch and Mesmer-Magnus (2010), the overwhelming finding across over 20 studies using a variety of study methods was that increased mental model overlap led to improved team performance. One method to improve overlap is through cross-training, whereby team members are trained on the other team positions to maximize shared knowledge. Marks et al. (2002) examined the link between team cross-training and team effectiveness. “Cross-training” was operationalized as team members learning about what each of the other team members was supposed to do in their role. Subject-matter experts identified ten task-related concepts that were essential for the team’s success. Each participant was asked to rate how relevant each of those ten concepts was to every role on the team. From these ratings, the researchers created a “similarity index” that quantified the degree of mental-model overlap among team members. The results showed that participants who received cross-training had a higher mental model overlap than those in the control group, and this greater overlap explained a substantial portion of the variance in both coordination and backup behaviors observed during performance (Marks et al., 2002). Most SMM research to date has emphasized overlap and has given comparatively little attention to complementarity. Yet the two mechanisms are not mutually exclusive; a team can benefit from a blend of shared and complementary knowledge, depending on the nature of the task and the coordination demands (Mohammed & Dumville, 2001). In practice, effective teaming often involves balancing the amount of common ground with the strategic distribution of unique expertise across members.

In addition to the methods of measuring the contents of individual mental models, there are also numerous ways to operationalize mental model similarity. In their review of SMM literature, DeChurch and Mesmer-Magnus (2010) described three dimensions of a mental model's operational definition: the way they are elicited, their structure, and their degree of convergence with other team members. Mental model elicitation involves asking participants to explain their understanding, experience, and explicit knowledge about something. This process captures the content of each team member's mental model, but not how that knowledge is organized. Mental model structure refers to how the knowledge in the model is organized, usually measured as associative networks. Finally, researchers can operationalize convergence - measuring how similar team members' mental models are to one another. Analyses such as pairwise comparisons and pathfinder analyses have been used to assess convergence, and Euclidean distance is used to evaluate the closeness of models between team members.

An illustration of convergence comes from Mathieu et al. (2000). Their empirical work linked teammates' mental models, team processes, and performance. Within SMM theory, the ability of a team to adapt successfully depends on how well members can anticipate what their teammates will do and when they will need to act. Shared mental models provide the knowledge base that enables each member to select actions that are consistent with, and coordinated to, the team's needs. Specifically, SMMs let members predict the information and resource requirements of their teammates, allowing them to act with an understanding of task demands and the impact of their actions on others.

Mathieu et al. (2000) recognized that there are likely multiple mental models co-existing among team members that represent elements of the task, and elements of the team. In their study, they measured team and task models separately and related them to indices of team processes and performance. A major purpose of their study was to measure mental model convergence among team members for both task and team models. They hypothesized that communication and strategy would be influenced positively by mental model similarity, and that this model similarity (or "convergence") would improve with experience.

In their study, participants worked in teams within a computer-generated simulation program for an F-16 fighter jet. This simulator was flexible, allowing researchers to script team-based scenarios in which teams flew a predefined route with waypoints, while trying to survive attacks and shoot down enemy planes. The study was conducted in three phases. During the first phase, participants received an overview of the task and watched an automated demo of the simulation. In the second phase, participants were guided through a structured hands-on training program. One player was trained to use a joystick and was responsible for flying the plane; the other used a computer keyboard to set airspeed, call up different weapons systems, and gather information. Both players were able to fire weapons. The training phase taught both players their individual task responsibilities as well as basic team processes. They also flew a practice mission together. In the third phase, players completed three performance "blocks" which involved answering a survey, conducting two missions, and receiving a summary of their performance in each mission that included waypoints reached and enemy planes shot down. Performance in each mission was evaluated using a points system, with points for survival, reaching waypoints, and shooting down enemy planes. Teams were presented with a strategic challenge in that these objectives conflicted with one another: shooting down enemy planes would not leave enough time to fly to each waypoint, and prioritizing survival limited the ability to earn points by shooting down enemy planes or reaching waypoints.

Team performance was operationalized in terms of team outcomes and team processes. Outcomes were measured using the points system described above. Team processes were rated independently by two observers watching videos of the missions. Each rater scored mission performance on 21 items intended to measure:

- Strategy formation and coordination (e.g., "To what extent did the team plan together to coordinate its efforts?").
- Cooperation (e.g., "To what extent did they cooperate well during the mission?").
- Communication (e.g., "To what extent was information about important events and situations shared within the team?").

Evaluators rated these items using a 5-point scale, from 1 indicating "not at all" to 5 "to a very great extent". Internal consistencies for both coordination and communication were high, though lower for cooperation.

Task and team mental models were analyzed for convergence among both team members. The authors defined a set of attributes for each model. These attributes were used to create matrices where individuals rated each attribute's relationship to every other attribute for that model using a point scale to indicate the direction (positive or negative) and strength of the relationship. To define attributes of the task model, researchers conducted a task analysis of the simulation using subject-matter experts and other technical resources. This analysis identified eight highly related but distinct attributes of the task mental model. Attributes of the team mental model were selected based on taxonomies of teamwork dimensions, especially as they apply to tasks such as the simulation used for this experiment. Attributes included those related to communication, coordination, and cooperation among team members, but also included things like roles, liking, and team spirit.

The matrices were completed by each team member before each of the blocks of mission performance. Task and team models were analyzed using a network analysis program called UNCINET, which provides an index of convergence between matrices (a strategy that captures "representation of emergence" discussed earlier). A correlation was also calculated between teammates' mental models at each time period, producing two indices for convergence for task and team mental models, which range from -1 (complete disagreement) to 1 (complete agreement).

Overall, team performance rose significantly between the first and second performance blocks, then tapered. However, while team process scores went up significantly with performance periods, both task and team models failed to show convergence over time and remained stable throughout the experiment. In other words, while team processes improved with time, mental model convergence among team members was not the reason for improved processes. Overall, team mental model similarity related significantly to team performance, mediated by team processes. Shared mental models did have a positive effect on team performance, although the models appeared to be stable even with additional experience. The ways that team members organize their individual team and task knowledge are predictors of performance, above and beyond simply having overlapping knowledge of the task.

Cooke et al (2013) describe the SMM theoretical perspective as one grounded in an information-processing model of individual cognition. In studying the mental models of individuals for similarity and convergence over time, proponents of SMM treat the collective knowledge of teammates as inputs to team processes. The result is an information processing model with the individual as the primary unit of analysis. Cooke and colleagues (2013) argue that a better way to predict team performance is by examining the team as the unit of analysis, rather than the individual. In the next section, we will review the other prominent model of team cognition – Interactive Team Cognition.

2.2 Interactive Team Cognition (ITC)

ITC was developed to provide explanatory power above and beyond shared mental models. Proponents of the theory argue that in settings that demand a high level of cognitive complexity, team members are unlikely to have uniform perspectives due to differing knowledge, skills, experience, and physical location in the environment. In these settings, teams are more likely to be a collection of unique perspectives that work together as a new entity within the task environment. As a result, team cognition is conceptualized as something that occurs at the level of the group, taking the form of explicit communication and coordination behaviors (Cooke et al., 2013). The ITC framework makes the team itself the cognitive system, and its members the components of that system, with cognition emerging through their interactions. An important difference between this and mental model complementarity is the locus of cognitive activity. Mental model complementarity studies team performance from an individual information processing standpoint. Critically, ITC rejects the information processing metaphor as the central theoretical paradigm, and replaces it with three premises (Cooke, 2013):

1. Team cognition is an activity, not a property or product,
2. Team cognition should be measured and studied at the team level, not the individual level, and
3. Team cognition is inextricably tied to context.

These premises differ from SMM theory in the sense that they focus on process (team interactions) instead of more static cognitive structures. Proponents of the ITC theory still recognize the contribution of knowledge overlap among team members. However, theoretically shared mental models would only facilitate early team performance. Later, as teamwork develops, ITC posits that interaction among team members has a greater contribution toward team success than shared knowledge (Cooke et al., 2013).

Interpretations from a body of research have led to the premises of ITC listed above (e.g., Cooke & Shope, 2004; Hackman & Katz, 2010; Gorman et al., 2010). For instance, several studies taking place in an Unmanned Aerial Vehicle (UAV) synthetic task environment (STE) have found that team performance is largely attributable to team coordination activities as opposed to shared knowledge. Cooke and Shope (2004) trained teams of three to simulate flying a drone to take photos for a reconnaissance mission. Teams involved three roles: pilot, photographer, and navigator. Each team member was trained to complete the tasks belonging to their role, then came together as a team to photograph a set of ground targets. Communication content and timing were recorded, as well as speaker and listener identities. Performance was assessed by the number of targets successfully photographed, route violations, and the amount of fuel and film used. A number of individual and team performance measures were also incorporated into the study. Over the course of several experiments, Cooke et al. (2013) describe the observation that improvements to team performance coincide with improvements in team interactions (including team process behaviors such as communication and coordination), but not with measures of shared knowledge.

Another study using the UAV testbed compared the effects of three types of training on skill retention and skill transfer in novel conditions (Gorman et al., 2010). These training approaches were procedural training, cross-training, and perturbation training. In perturbation training, teams are presented with disruptions during the task that force them to coordinate to overcome the disruption. Gorman and colleagues found that perturbation-trained teams demonstrated better performance under circumstances that required adapting to novel conditions (Gorman et al., 2010). In addition, Cooke, Gorman, Duran and Taylor (2007) found that teams with prior command and control experience, but no prior UAV experience, had better performance within the task environment, while demonstrating no increased shared knowledge. The researchers attributed the advantage of those teams to the experience they had coordinating with one another in command-and-control contexts. These studies collectively show that the interaction among team members is a superior predictor of team performance compared to shared knowledge. Moreover, Cooke states that team cognition is an emergent dynamic activity that involves the team as a whole interacting with an uncertain and evolving task environment (Cooke et al., 2013). This is the basis for the premise that team cognition is an activity, rather than a property or product of individual team members. This cognitive activity can be observed using metrics focused on team interaction. Examples include communication flow (such as who talks to whom) (Cooke et al., 2013). Importantly, Cooke and colleagues (2013) assert that because of the necessity of measuring team coordination and communication, which involves more than one individual, team cognition cannot be studied at the level of the individual. For this reason, ITC studies team cognition at the team level of analysis (the second premise of ITC) (Cooke et al., 2013).

The third premise of ITC posits that teamwork is inextricably tied to context. ITC theory has theoretical roots in perspectives that recognize the interaction of the cognitive system with the external environment (Cooke et al., 2013). An ITC theory of team cognition must account for the ubiquitous surrounding context of team member interactions and a dynamic environment (Cooke & Gorman, 2009)

ITC also assumes that other models (like SMM) conflate schematic overlap among team members with the study of mental models of team members. In other words, to study mental models at the individual level is to study the organization of knowledge within the individual, but not the specific dimensions of an ongoing situation that may or may not be guiding perception of that situation.

2.3 Bridging Individual Cognition to Team Performance

ITC offers an important perspective by focusing on dynamic activity. While SMM theory has defined team cognition in terms of similarity or complementarity of largely static knowledge structures among individuals, this perspective fails to capture the dynamic changes in the environment and interactions among team members (Cooke et al., 2013). However, by limiting its scope to communication and coordination behavior, ITC places less importance on modeling what is happening within the individual's ongoing mental representation and more on being able to predict the team's outcome. This is evident in the following quote from Cooke's 2013 review:

"Specifically, ITC suggests that team cognition needs to be measured in the context of the unfolding task with the focus on activity occurring at the team level, not relatively static (ex situ) knowledge representations."

But individual knowledge representations are not limited to the static models the authors reference here. Rather, there are existing models for individual cognition that reflect many of the same important circumstantial and environmental contributions that ITC highlights – namely, that team performance is context dependent, and that it can be described actively as a process rather than an end product. Furthermore, viewing individual cognition from a perspective that readily incorporates context-dependency has utility for designing digital twins, or otherwise programming AI agents that are useful in a human/AI hybrid teaming environment. A model of individual cognition also lends itself to the development of high-fidelity measures of performance. Additionally, these models should, in theory, be able to predict the team-level behaviors that are central to ITC's focus, namely communication and coordination behaviors among teammates. In the following section, we will explore a model of cognition that embodies these qualities at the individual level of analysis.

3 Event Cognition

Event cognition is the study of how people perceive, conceive of, talk about, and recall events. The way that cognitive systems form and update representations of events is a central focus of this research (Radvansky & Zacks, 2017). Event cognition has emerged as a theoretical framework for a breadth of domains, from the study of reading comprehension (Zwaan et al., 1998) to the comprehension of narrative film (Kurby et al., 2014), and even the comprehension of first-person events (Magliano, 2014).

Critical to event cognition is the concept of the "situation model". The situation model is an integrated mental representation of a state of affairs (Zwaan et al., 1998) and is a concept that has been applied to the study of situation awareness (Endsley, 2000). Situation awareness can be described as perceiving critical factors in the environment (Level 1 SA), understanding what those factors mean, particularly when integrated together in relation to the person's goals (Level 2 SA), and at the highest level, understanding of what will happen within the system in the near future (Level 3 SA) (Endsley, 2018). Endsley describes the components of perception, comprehension, and projection in the context of the ongoing situation model. Consider the following passage from Endsley's 2000 review of theoretical underpinnings of situation awareness:

"In this sense, SA, or the situation model, is the current state of the mental model. For example, one can have a mental model of a car engine in general, but the situation model (SA) is the state it is believed to be in now (e.g., it has a given RPM, it has a given level of fuel, it is at a given temperature, etc.). This situational model captures not only the person's representation of the various parameters of the system, but also a representation of how they relate in terms of system form and function to create a meaningful synthesis -- a gestalt comprehension of the system state."

Figure 1 comes from the same review, and depicts the relationships between the situation model, mental model, elements of the environment, and active goals. In this diagram, the mental model is a stable feature of long-term memory that represents components of the environment, their form, function, and purpose, and the interrelation among those components. This model directs attention and perception

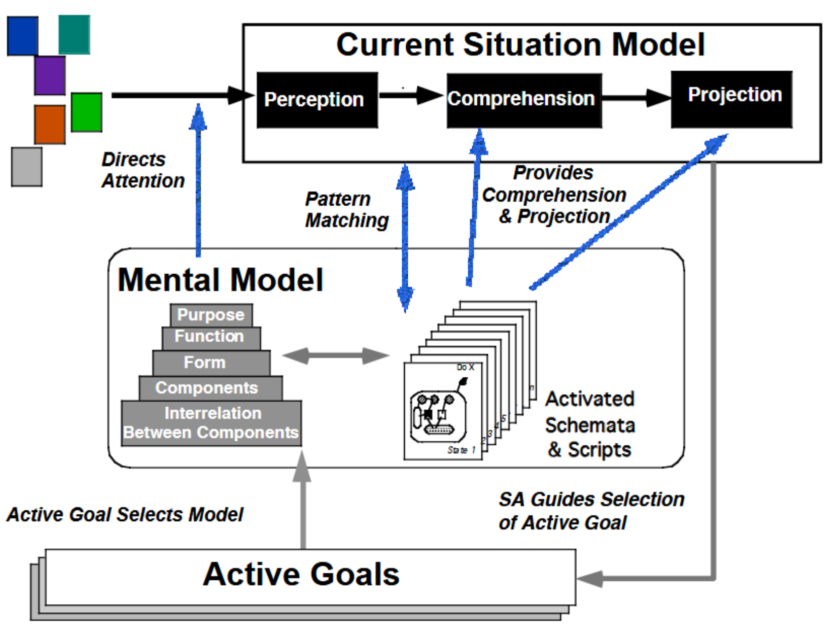

Figure 1. Situation Model Diagram from Endsley, 2000.

within the environment. It also activates schemas and scripts that inform comprehension and projection, which in turn influence pattern matching in making sense of the environment. The notion of a "current situation model" is useful for understanding situation awareness in the sense that it provides an artifact to target and study. For example, Endsley has described the use of SA measures as a way to take a "snapshot" of the mental model. Further, she posits that these measures could be used as inputs for a computer-based situation model. Endsley's model for situation awareness has run into some familiar roadblocks, one being the process vs. product conflict we have described earlier in this report. A 1995 criticism of Endsley's prior work required distinguishing product and process. She therefore refers to the process as situational assessment and the product as situational awareness. However, she clarifies that situational awareness is dynamic and not merely a knowledge state (Endsley 2015, misconceptions).

The study of event cognition has been utilized for research in other domains as well. For instance, reading comprehension has made use of frameworks such as the "structure building framework" (Gernsbacher, 2014) and the "event indexing model". These models have supported event cognition research for reading text, for watching narrative film, and for understanding first-person events. They identify specific dimensions of events as the basis for updating the perceiver's situation model, including:

- Agent locations
- Agent goals
- Intentional entities
- Other relevant "entities" such as animals or objects
- Abstract concepts such as temporal and causal relationships

A shift along any of these dimensions indicates that the event model is no longer relevant and must be modified or reconstructed.

Event Segmentation Theory (EST) goes further in offering an explicit account of how situation models for events are updated. Whenever a change along a meaningful dimension of activity is detected, a boundary within the event is perceived, and the event model is updated (Reynolds, Zacks & Braver, 2007). In EST, individuals form event models in working memory that guide ongoing perception, facilitating sense-making and predictions of what will occur in the environment (sometimes referred to as "anticipatory processes"). When those anticipatory processes are correct, the current event model stays active. When they fail, the event model is updated. When this happens, the individual is more sensitive to new information in the environment, and overall awareness is heightened. After updating, a new event model emerges, and the cycle starts over again. This chain of events resulting in the new event model is perceived as a boundary in the event (Zacks, Speer, & Reynolds, 2009).

Event boundaries often occur when the environment changes, as those changes tend to introduce unpredictability. These can be perceptual changes, such as changes in movement, color, and sound, but can also be conceptual, such as changes to characters in a story, the status of their goals, and other cause and effect relationships (Zacks et al., 2009). Thus, event boundaries usually occur when an individual perceives changes in physical situational attributes like location, the passage of time, interaction with objects, interaction with other agents, and with conceptual situational attributes such as changes in goals, causality, etc.

Pettijohn and Radvansky (2016) conducted research on the "location updating effect" – the phenomenon in which information becomes less accessible in memory when it is accompanied by a shift in space. The paradigm for studying this effect is one where participants navigate virtual environments, picking up objects and putting them down. A critical manipulation is whether the participant carries a particular object across a virtual room, or over a threshold into another room in the virtual space. In their study, Pettijohn and Radvansky had participants navigate a virtual space, either carrying virtual items across a room and putting them down or carrying them to another room and setting them down. Probes appeared at the center of the screen that required a yes or no response as to whether they had encountered that item. Half of these probes were items that participants had interacted with in the virtual space, and half were not. Error rate, response time, and travel time were all measured. Results from their study revealed that the error rate was significantly higher for items in the shift condition (where participants carried them from one room to the next) compared to the no shift condition. Corroborating this finding, longer response times were also found for the shift condition. Importantly, there was no difference between conditions in the amount of virtual space or time experienced by the participant. The only difference between conditions was whether the object in question was transported between two virtual rooms. Interestingly, this effect is not merely perceptual. The researchers manipulated visual continuity between the rooms by creating "glass" walls so that perception of the space was not interrupted when entering a new room. Despite this, walking through the virtual doorways still impacted memory for items in the environment. It was only when perceptual change was interpreted as a new region in space that a boundary was perceived and memory was affected. These findings suggest that memory for the object diminishes when it is transported to another room (the shift condition) compared with when it is left in the same room when it is set down (the non-shift condition). The explanation for this finding is that a new event model is created for the second room in the shift condition. When the participant in the shift condition receives the recognition probe for the item in question, two competing event representations are invoked, leading to interference with the response. This indicates that top-down processes (such as preconceived notions of what constitutes a "room") also exert significant influence on the perception of event boundaries.

Zachs, Speer and Swallow (2010) have also used EST as a framework for studying event cognition. They had participants watch a film called "The Red Balloon" in which a boy interacts with a seemingly sentient balloon, and complete segmentation tasks. In one task, while watching the film, participants were asked to press a button to identify the largest natural and meaningful unit of activity they perceived. In the other, they were asked to press a button to identify the smallest units of natural and meaningful activity. The Red Balloon had been previously scored for six types of changes, with timestamps at changes for each type identified by two independent raters. These changes were

- Spatial Changes: coded when a character changed direction within a scene or when the point of view of the camera changed location.
- Temporal Changes: coded whenever the frame after a cut was temporarily discontinuous from the prior frame.
- Object Changes: coded whenever the nature of a character's interaction with an object changed, such as picking it up or using it in a new way.
- Character Changes: coded when the focus of action was an animate character or characters, and this focus was different from that of the preceding frame. Character interactions were defined as changes in physical or abstract interactions between characters (talking, gesturing, etc.).
- Causal Changes: coded whenever the activity in a frame could not be described as directly caused by something in the previous frame.
- Goal Changes: coded whenever a character performed an action associated with a different goal from the goal in the previous frame.

Event boundaries were associated with changes in all dimensions in the list above and were predictive of both coarse and fine-grained segmentation. In other words, the likelihood that a participant detected an event boundary was significantly increased by a change in one or more of those dimensions.

A common theme that we can derive from the collective body of literature on event cognition is that the situation model that drives attention is influenced both by bottom-up perceptual processes and existing mental models, or “top-down” processes. These top-down processes are of particular interest for a teaming context. Just as knowledge of narrative structure, such as understanding character interactions and goals, drives narrative film comprehension, it is likely that knowledge of teaming and of team members guides team efforts. If so, what are the dimensions of mental models that guide an individual’s effort within a team? Do they change based on the type of team an individual is engaged with?

Some of the research that begins to explore this question is that of team situation awareness. Kaber & Endsley (1998) posit that team members have “sub goals” that dictate the focus of their attention. However, the “weak link” in overall team SA is someone who needs a piece of information but cannot acquire it. To prevent this, teams require a certain amount of awareness of teammates and their goals as well, depending on the level of interdependency that is required for their success. If different team members have different assessments of shared SA requirements, it could affect their coordination, leading to errors. This implies that as team interdependency increases, so does the necessity of teammates to represent teammate actions, and therefore anticipate potential needs, in their situation model. This would not be required for a similar task in which teammates operate more independently. Consistent with this, Schelble et al. (2025) conducted a study looking at how participants in hypothetical hybrid teams perceived information sharing from synthetic agents and found that design features of agents that impacted the general situation awareness of the team had the largest positive effect on perceived team cognition. While this is an interesting starting point, Schelble’s work was focused on the impression of team members in a hypothetical scenario, and did not test whether this perception of enhanced team cognition results in superior team performance in an operational environment.

4 Expanding on Team Cognitive Models

Why might understanding mental models that guide moment-to-moment processing in a teaming context matter in the larger scope of teaming research? To answer this question, we revisit the ongoing research goals surrounding hybrid teams. The National Academies of Science, Engineering, and Medicine (NASEM)[2] report identified nine focus areas within the domain of Human-AI-teaming. Among those focus areas were "human-AI team interaction", "trusting AI teammates", "human-AI teaming methods and models", and "situation awareness in human-AI teams" (NASEM, 2021). These focus areas emphasize an understanding of teammates (human and synthetic), and of the task environment. While ITC takes the perspective that the individual is not an appropriate unit of analysis for a team, a theoretical perspective that could effectively explain an individual's cognition in a team setting has potential utility for enhancing hybrid teaming research through the design of digital twins or programming AI agents that are useful in a teaming environment. In order to anticipate human teammate needs, a synthetic agent must be able to represent the needs of an individual teammate and respond meaningfully.

Additionally, developing an understanding of team cognition at the individual level can facilitate higher fidelity measures of behavior, thus enhancing the downstream predictive ability of teaming models. While card sorting tasks or diagramming are useful measures, they are relatively low-fidelity cognitive measures. Lower-level cognitive behaviors (event segmentation, timed response, eye tracking) and physiological measures (pupil dilation, galvanic skin response, or GSR), when predicted by a robust model of cognition, offer higher fidelity measurements of cognitive processes. Higher fidelity measures facilitate fine-tuning of cognitive models, which can better inform the design and behavior of synthetic agents in a hybrid teaming environment.

The first step toward establishing a cognitive model that is predictive of team performance is to expand the situation awareness and event comprehension literature to incorporate teaming. In the same way that the narrative comprehension literature identifies dimensions of events that are predictive of shifts in a reader's situation model, this research can identify and assess the predictive value of dimensions of team-related task environments for team environments. The goal of this research would be to demonstrate how the emerging situation model in the individual is influenced by the team operational environment and reciprocally influences behavior within that environment. A simple starting point for this research may be the hypothesis that in an interdependent team workflow setting, teammates are more active constituents of an individual's mental representation (i.e., are at a state of higher neural activation) than they are in a pooled or sequential workflow setting, where team member performance is independent.

The situation model from Endsley's SA work describes the chain of perception-comprehension-projection as a key part of the sense-making act within teams. This tripartite link is mostly studied at the individual level within teams, capturing the person's representation and comprehension of a system. The proposed expansion is not meant to replace this paradigm, but to extend its functions into a teaming scenario. This work re-emphasizes the importance of the individual as the unit of analysis, because every individual represents a potential "weak point" that could negatively impact the communication and coordination behaviors critical to successful team performance.

[2] Described and referenced in prior technical reports in this series.

4.1 Data Collection Methods

There are a number of ways to devise a study to explore the relative activation of teammates within an individual's situation model. In this section, we will explore a few research paradigms and related measurement methods and tools for evaluating the contents of an individual's situation model.

4.1.1 SAGAT: Endsley

The Situation Awareness Global Assessment Technique (SAGAT) is a global tool developed to assess SA across all of its elements based on a comprehensive assessment of operator SA requirements (Endsley, 1995). Using SAGAT, a simulation employing a system of interest is frozen at randomly selected times, and operators are queried for their perceptions of the situation at that time. The system displays are blanked, and the simulation is suspended while subjects quickly answer questions about their current perceptions of the situation. As a global measure, SAGAT includes queries about all operator SA requirements, including Level 1 (perception of data), Level 2 (comprehension of meaning), and Level 3 (projection of the near future) components. This includes a consideration of system functioning and status, as well as relevant features of the external environment. SAGAT queries allow for detailed information about subject SA to be collected on an element-by-element basis that can be evaluated against reality, thus providing an objective assessment of operator SA. This type of assessment is a direct measure of SA — it taps into the operator's perceptions rather than infers them from behaviors that may be influenced by many other factors besides SA. Furthermore, it does not require subjects or observers to make judgments about situation knowledge on the basis of incomplete information, as subjective assessments do. By collecting samples of SA data in this manner, situation perceptions can be collected immediately (while fresh in the operators' minds), reducing numerous problems incurred when collecting data on mental events after the fact, without incurring intrusiveness problems associated with online questioning. Multiple "snapshots" of operators' SA can be acquired in this way, providing an index of the quality of SA provided by a particular design. By including queries across the full spectrum of an operator's SA requirements, this approach minimizes possible biasing of attention, as subjects cannot prepare for the queries in advance since they could be queried over almost every aspect of the situation to which they would normally attend. The method is not without some costs, however, as a detailed analysis of SA requirements is required in order to develop the battery of queries to be administered.

4.1.2 Verbal Protocol Analysis

Verbal Protocol Analysis (VPA) (Green, 1995; Lee et al., 2012) is a research method used to study human cognition during task performance by analyzing what people say aloud while they are performing a task or immediately after completing it. People's verbalizations provide observable traces of otherwise unobservable cognitive processes. In VPA, participants perform a task (e.g., problem solving, decision making, monitoring a system). They are asked to think aloud during the task (concurrent protocol) or explain what they were thinking after the task (retrospective protocol). Speech is audio/video recorded and transcribed. Transcripts are segmented and coded according to a predefined scheme (e.g., goals, cues noticed, interpretations, decisions). Patterns in the verbal data are used to infer cognitive states or processes. Importantly, VPA focuses on processes, not just outcomes.

1. Concurrent VPA is closer to actual cognition. However, it is prone to interfere with task performance.
2. In retrospective VPA, the participant explains their reasoning after task completion (often aided by replay). While there is less interference, it relies on memory and reconstruction.

4.1.3 Situation Awareness Rating Technique (SART)

The Situation Awareness Rating Technique (SART) (Taylor, 1989) is a widely used subjective workload and situation awareness measurement tool. Developed originally for the aviation industry, it relies on the operator's self-assessment to gauge how well they understand their environment. SART is based on the 3-D model, which aggregates ten specific dimensions into three primary domains. It is typically administered via a post-task questionnaire where participants rate their experience on a 7-point Likert scale.

The Three Pillars of SART are:

1. Demands on Attentional Resources: How much is the situation "asking" of the person? (Complexity, variability, and instability of the situation).
2. Supply of Attentional Resources: How much "energy" or focus can the person give? (Arousal, concentration, and spare mental capacity).
3. Understanding of the Situation: How much does the person actually "get" what's happening? (Information quantity, quality, and familiarity).

The final SA score is calculated using the formula: SA = Understanding - (Demands - Supply). Teaming research views team performance as a series of cognitive "episodes"—shifting from individual processing to collective action. To determine if SART is a viable tool for this model, we have to look at how well a subjective, individual metric fits a dynamic, team-based framework. While SART is a low-cost method that is easy to administer between "episodes" in any team cognition task, it relies on memory, which can be distorted at the end of an episode (not least due to various types of cognitive biases). SART is a good tool to measure SA on a supply-demand scale, but it cannot capture the distributed nature of SA between various members of the team. This is because Team SA is more than just the sum of individual SART scores. Therefore, the viability of SART as a research paradigm for our purposes is questionable.

4.1.4 Subjective Workload Dominance (SWORD)

While SART focuses on "how much do I know," Subject Workload Dominance (SWORD) focuses on relative workload—specifically, which task was harder than the other. Developed by Michael Vidulich, SWORD uses a pro-comparison approach based on the Analytic Hierarchy Process (AHP) (Vidulich, 1989). Instead of rating a task on an absolute scale of 1–10, the participant compares two different tasks or "episodes" and decides which one was more demanding and by how much. The SWORD methodology includes these three steps:

1. Pairwise Comparison: If you performed Task A and Task B, you are asked: "Which task had a higher workload?"
2. Magnitude Rating: You then indicate the strength of that preference on a scale (e.g., Task A was "moderately more demanding" than Task B).
3. Judgment Matrix: These comparisons are plugged into a square matrix to calculate a weighted workload score for every task performed.

Due to the process-oriented nature of team research, we might be interested in studying the flow of cognitive episodes. SWORD may be suited for this because it is suited to compare one episode against another. With event cognition's focus on transitions between episodes, SWORD-like tools will be beneficial in identifying exactly when a team's cognitive load spiked or troughed compared to prior states. Research shows that humans are generally better at making relative judgments, rather than absolute judgments. To that end, SWORD allows team members to look back and rank the difficulty of various phases. However, like SART, this tool also relies on episodic memory and is hence prone to cognitive biases that accompany human recollections. Further, SWORD becomes computationally expensive very quickly

(N^2 computations balloon beyond, say, 10 episodes). This can lead to rater fatigue and might not be scalable for long-horizon tasks. Further, while SWORD tells you about the relative magnitude between two episodes, it does not shed light on the reasons. SART is somewhat better at diagnosing the “why”.

4.2 Research Directions

Earlier in this section, we outlined a simple hypothesis about the relative activation of team members in the situation model. Using that as a starting point, we can also expand this area of research to include other inquiries. Each of these questions represents more starting points to expand this area of research. With forthcoming research, we will be able to better formulate these questions and unravel other avenues of study. Additional research questions include:

- Which (if any) other dimensions are critical to guiding situation models in teaming task environments?
- How do types of interdependencies (pooled, sequential, reciprocal) (Thompson, 2003) in team workflows influence situation models?
    - For a fixed task (characterized by fixed goals, objectives, and scoring mechanisms), how do variations in interdependence conditions impact different teammates’ abilities to perceive critical events within a team?
- How do situation models in a teaming scenario inform the long-term stable mental models of the team?
    - This question seeks to address the impact of individual events in the team, and how their perception changes long-term mental models of the team, and explore whether they have long-term influence over how teams and team tasks are understood.
- How do other types of task-related interdependencies (such as action, goal, resources) (Golpayegani et al., 2019) between teammates impact the development of the situation model within a teaming context? Are there threshold levels for interdependence at which these impacts are observed?
- Do elements of the situation model that occur at the individual level predict activity at the team level defined by interactive team cognition?
- Are there additional individual and group-level units of measurement and analyses for measuring dimensions of situation models that evolve in a teaming environment? How do we reconcile the individual and group-level units of analysis?

This Page Intentionally Blank

References

Allen, R. B. (1997). Mental models and user models. In *Handbook of Human-Computer Interaction* (pp. 49-63). North-Holland.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and Group Decision Making: Current issues* (pp. 221–246). Lawrence Erlbaum Associates, Inc.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *22*(2), 195-202.

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied*, *13*(3), 146.

Cooke, N. J., & Gorman, J. C. (2009). Interaction-based measures of cognitive systems. *Journal of cognitive engineering and decision making*, *3*(1), 27-46.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255-285.

Cooke, N. J., & Shope, S. M. (2004). Synthetic task environments for teams: CERTT's UAV-STE. In *Handbook of human factors and ergonomics methods* (pp. 476-483). CRC Press.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 1.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (2000, July). Situation models: An avenue to the modeling of mental models. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 44, No. 1, pp. 61-64). Sage CA: Los Angeles, CA: SAGE Publications.

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of cognitive Engineering and Decision making*, *9*(1), 4-32.

Endsley, M. R. (2018). Automation and situation awareness. In *Automation and human performance* (pp. 163-181). CRC Press.

Gernsbacher, M. A. (2014). The structure-building framework: What it is, what it might also be, and why 1. In *Models of understanding text* (pp. 289-311). Psychology Press.

Golpayegani, F., Dusparic, I., & Clarke, S. (2019). Using social dependence to enable neighbourly behaviour in open multi-agent systems. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–31.

Gorman, J. C., Cooke, N. J., & Amazeen, P. G. (2010). Training adaptive teams. *Human Factors*, *52*(2), 295-307.

Green, A. (1995). Verbal protocol analysis. *The Psychologist, 8*(3), 126–129.

Gupta, N., & Hollingshead, A. B. (2010). Differentiated versus integrated transactive memory effectiveness: It depends on the task. *Group Dynamics: Theory, Research, and Practice*, *14*(4), 384.

Hackman, J. R., & Katz, N. (2010). Group behavior and performance. *Handbook of social psychology*, *2*(1208-1251), 2.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71-115.

Johnson-Laird, P. N. (2005). Mental models and thought. *The Cambridge Handbook of Thinking and Reasoning*, 185-208.

Kaber, D. B., & Endsley, M. R. (1998). Team situation awareness for process control safety and performance. *Process safety progress*, *17*(1), 43-48.

Kurby, C. A., Asiala, L. K., & Mills, S. R. (2014). Aging and the segmentation of narrative film. *Aging, Neuropsychology, and Cognition*, *21*(4), 444-463.

Lee, S. W., Park, J., ryum Kim, A., & Seong, P. H. (2012). Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis. Annals of Nuclear Energy, 43, 167-175.

Magliano, J. P., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology*, *26*(6), 649-661.

Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, *87*(1), 3.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of applied psychology*, *85*(2), 273.

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *22*(2), 89-106.

National Academies of Sciences, Engineering, and Medicine. (2021). Human-AI Teaming: State-of-the-Art and Research Needs.

Pettijohn, K. A., & Radvansky, G. A. (2016). Walking through doorways causes forgetting: Event structure or updating disruption? *Quarterly Journal of Experimental Psychology*, *69*(11), 2119-2129.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, *17*, 133-140.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological bulletin*, *133*(2), 273.

Schelble, B. G., Flathmann, C., Macdonald, J. P., Knijnenburg, B., Brady, C., & McNeese, N. J. (2025). Modeling perceived information needs in human-AI teams: improving AI teammate utility and driving team cognition. *Behaviour & Information Technology*, *44*(9), 2069-2092.

Taylor R.M. (1989) *Situational Awareness Rating Technique (SART): The Developement Of A Tool For Aircrew Systems Design*. In Proceeding of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478. Seuilly-sur Seine: NATO AGARD.

Terrien, E., Huet, B., Iachkine, P., & Saury, J. (2020). Coordination between crew members on flying multihulls: a case study on a Nacra 17. *Journal of Sports Science & Medicine*, *19*(2), 298.

Thompson, J.D. (2003). Organizations in Action: Social Science Bases of Administrative Theory (1st ed.). Routledge.

Vidulich, M. A. (1989, October). The use of judgment matrices in subjective workload assessment: The subjective workload dominance (SWORD) technique. In Proceedings of the Human Factors Society Annual Meeting (Vol. 33, No. 20, pp. 1406-1410). Sage CA: Los Angeles, CA: SAGE Publications

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). New York, NY: Springer New York.

Wegner, D. M. (1995). A computer network model of human transactive memory. *Social cognition*, *13*(3), 319-339.

Yoo, Y., & Kanawattanachai, P. (2001). Developments of transactive memory systems and collective mind in virtual teams. *The International Journal of Organizational Analysis*, *9*(2), 187-208.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*(2), 307.

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in human neuroscience*, *4*, 168.

Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific studies of reading*, *2*(3), 199-220.